DOCUMENT RESUME

ED 069 713 TM 002 165

AUTHOR	Wilson, Clark L.; And Others
TITLE	A Manual for Use in the Preparation and Administration of Practical Performance Tests.
INSTITUTION	Office of Naval Research, Washington, D.C. Psychological Sciences Div.
REPORT NO	NAVPERS-91961
PUB DATE	71
NOTE	55p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Job Skills; *Manuals; *Military Personnel; *Performance Tests; Scores; Scoring; Skill Analysis; Task Analysis; Task Performance; *Test Construction; *Testing
IDENTIFIERS	Navy

ABSTRACT

The purpose of this manual is to help Navy division officers, petty officers, and instructors build, use, and understand practical performance tests. Methods for testing men's ability to perform practical jobs are described in order to determine what they can actually do in meeting the requirements of their ratings, billets, or school training programs. Chapters are included on: (1) where and how to use performance tests advantageously, (2) deciding what to use for test tasks and an overview of the various types of performance tests, (3) "final product" performance tests, (4) "checklist" performance tests, (5) the construction and use of "examinee recorded" performance tests, and (6) how tests should be given and what to do with the test scores that have been obtained. The test construction chapters each contain an explanation of the type of test task to use, recording procedure, scoring methods, and suggestions on writing instructions for the examiner and examinees. (Author/JS)

ED 069713

A MANUAL FOR USE IN THE PREPARATION AND ADMINISTRATION OF PRACTICAL PERFORMANCE TESTS

NAVPERS 91961

1

ED 069713

A Manual for Use in the Preparation and Administration of

PRACTICAL PERFORMANCE TESTS

Prepared for

Personnel Analysis Division
Bureau of Naval Personnel

and

Psychological Sciences Division
Office of Naval Research

by

CLARK L. WILSON, Ph.D.
ROBERT R. MACKIE, Ph.D.
DONALD N. BUCKNER

Management and Marketing
Research Corporation

Contract Nonr 876(00)

ARTHUR I. SIEGEL, Ph.D.
DOUGLAS COURTNEY, Ed.D.

Institute for Research in
Human Relations

Contract Nonr 872(00)

Foreword

The purpose of this manual is to help division officers, leading petty officers, and instructors build, use, and understand practical performance tests. The manual describes methods for testing men's ability to perform practical jobs, in order to determine what they can **actually do** in meeting the requirements of their ratings, billets, or school training programs.

In writing this manual, it was our aim to present in a simple way the steps to be followed in constructing performance tests, with the hope that good performance tests will be made and used for these purposes:

1. To help assign new men to billets in such a way as to achieve the best utilization of personnel.
2. To help determine when a man can actually do the job for which he is being trained.
3. To determine who is best prepared to handle special job codes.
4. To determine who is qualified for advancement in rating.
5. To measure what in-service trainees or school graduates have achieved.
6. To measure trainee achievement within phases or units of school curricula.
7. To locate strengths and weaknesses in training programs.
8. To determine the qualifications of men entering training schools.

WHAT IS A PERFORMANCE TEST?

A performance test is a sample work situation in which the man being tested performs a practical task. The test requires him to demonstrate how well he has mastered one or more of the skills required for the actual performance of his job. A performance test may require the examinee to perform one of the tasks which he performs in his everyday routine, or it may involve a specially designed task which requires him to demonstrate a skill essential to his rate, billet, or school situation.

Performance tests differ from written or identification tests in that, in the performance test, the examinee is actually required to do a job. Written and identification tests are designed to measure how much a man **knows about** his job and the equipment he works with, but do not reveal how he may **perform** or use his tools when confronted with a job to actually do.

It is not intended that performance tests replace other methods of evaluation such as written tests, identification tests, or quarterly marks. They do, however, provide an important addition to these other methods. They should be used as part of the whole scheme for evaluating a man, since they provide information that cannot be obtained in other ways.

HOW TO USE THIS MANUAL

Chapter I of this manual tells where and how performance tests can be used to advantage. The chapters which follow it explain in detail how performance tests are built.

Chapter II, "Test Tasks and Types of Tests," explains how to decide what to use for test tasks and gives an overview of the various types of performance tests. This chapter will give you an idea of the type of test that will fit your needs best.

Chapter III covers "Final Product" performance tests. These are useful when you want to use performance tests to test men on their ability to turn out an end product which adheres to prescribed dimensions or which is free from errors or defects.

Chapter IV deals with "Check List" performance tests which are used when an examiner is available to actually watch the examinee perform a complex task. The latter part of this chapter covers the combination of final product and check list tests.

Chapter V explains the construction and use of "Examinee Recorded" performance tests, which you will want to use when you have a large number of men to test and only a few examiners to supervise their performance. They are especially valuable in measuring practical achievement in service schools.

Chapter VI describes how tests should be given and what to do with the test scores once they have been obtained.

The three chapters on test construction each contain an explanation of the type of test task to use, the procedure for recording performance, the methods of scoring performance, and suggestions on writing instructions for the examiner and examinees.

CHAPTER I

Advantages and Uses of Performance Tests

Probably the best way to give the reader an understanding of performance testing is to show, in some detail, the advantages of this method of testing and to show its many uses.

ADVANTAGES OF PERFORMANCE TESTS

1. **They Are Work Samples.** Performance tests are the best means available for measuring men's ability to **perform** their jobs. Other testing methods, such as written examinations, can tell you how much men know **about** their jobs or how familiar they are with equipment, but performance testing alone can tell you how well men can actually **do** the jobs required by their rates.

It is not necessary to have a man perform all the jobs required by his rate to get a good indication of his qualifications. Instead, his performance of a group of representative tasks (a work sample) will tell you how well he has mastered the basic skills of his rate, and thus give you an over-all picture of his ability to do his job.

2. **They Are Non-verbal.** In those rates which call for mechanical or technical skills, a man often learns more by **doing** than he does by reading or listening to lectures. The emphasis in these rates is less upon verbal ability and more upon a man's ability to use his knowledge effectively in performing practical

WORK! NOT WORDS

jobs. As a result, a man who is not able to express himself too well verbally can still learn the skills of his rate and perform his job with the best of them. That is, learning to fulfill the duties of a mechanical or technical rate does not require a man to be highly skilled in the use of language.

On the other hand, it is well known that some men can "talk" a good job but not perform it well when the time actually comes to do it. It is most important in such cases that the measurement of a man's mastery of the basic skills be based upon his ability to perform the real job rather than his ability to tell you how much he knows about it.

Therefore, because performance tests stress the **doing** of a job rather than the **answering of questions** concerning a job, they are useful far beyond the point where other testing methods leave off.

3. **They Are Objective.** Because every man who takes the same performance test performs a specific task under exactly the same conditions, and because every man who takes the same test is scored in exactly the same way no matter who scores him, performance tests give a fair, objective measure of the men's mastery of the skills involved in doing the specific test task. If performance tests are used, it is not necessary to rely on a division officer's or leading petty officer's **opinion** of what a man can do. Instead, you have an objective measure of the man's actual ability to perform. By setting up these standard conditions of measurement, the possibility is reduced of a man being rated high or low because of the personal opinions of his superior. Opinions of a man's technical proficiency are often derived, at least partially, in terms of the man's personality.

It is not intended, here, to indicate that technical proficiency is everything to the exclusion of a man's ability to be a leader and get along with people. Performance tests are measures of technical proficiency to be used in addition to measures of other characteristics which may be equally important. They are the best measures of technical proficiency and, as such, do not pretend to measure other characteristics such as leadership.

4. **They Are Easy to Administer.** Personnel in Navy schools and operating u cs helped in the preparation of this manual, and it was the finding of leading petty officers and school instructors that performance tests not only were effective but also were easy to use. This refers to the fact that the techniques of testing were relatively easy to master and the kinds of tests that gave the most objective measures of performance made the least demands on the examiners' time. In some instances, it was even possible to conduct a performance test in a regular watch standing situation.

USES OF PERFORMANCE TESTS

The purposes for which performance tests are built will differ widely from one organization to another. Their uses will be different in air squadrons, aboard ships, and in training schools. But their wide applicability in all these situations speaks for their value. You, as the user of performance tests, are in the best position to know how and where performance tests can help you most.

Some of the more important uses of performance tests are listed below:

1. **They Help You Select and Assign Men.** If you are receiving a new man or draft of men into your organization, you will want to assign them to billets according to their ability. Performance tests can tell

THEY HELP SELECT MEN

you whether a man is qualified to stand a given watch or perform a given duty. Or, they can tell you which of several men is the **best** qualified to do the task.

2. **They Tell You When a Man Is Ready to Perform a Job.** Both aboard ship and in the schools, it is necessary to know when a man has learned enough to be able to perform the tasks he is being taught to do. In the shipboard situation you may want to know if he's ready to perform the duties required for standing a watch, or you may want to know if he can repair equipment or adjust it well enough to perform the task by himself. Performance testing can give you answers to these questions and tell you when a man is ready to perform a given job or to stand a watch.

3. **They Tell You Who Is Best Prepared for Special Job Codes.** In many Navy units, afloat and ashore, there are certain special Navy job codes assigned to men who are capable of performing specialized duties. In these situations, performance tests can be used to indicate which of the men from a group of men or how many from a group are prepared to do these jobs and therefore should be assigned the special job codes.

4. **They Tell You When a Man Is Ready for Advancement in Rating.** Advancement in rating is based, among other things, on a man's score on a competitive written examination and also upon his ability to perform the practical factors of his rate. In the past the leading petty officer or division officer often indicated, from his experience in watching the man perform his daily duties aboard the ship, whether or not he thought the man had the skill necessary for advancement. Performance testing offers an objective, standard method with which qualifications for advancement in rating can be measured and upon which recommendations for advancement in rating can be based. Performance tests can tell you which men are most qualified for advancement or which have reached the minimum level of qualification for advancement. In this way performance tests can be used in the same manner written tests are now used. They would not be a substitute for written tests, but a supplement, measuring practical ability in the same way written tests measure job knowledge.

5. **They Can Help You Establish the Qualifications of School Graduates.** It is a common practice in the Navy to indicate how well a man has done in a school by reporting his class standing. Usually, this class standing is based on an average of the written tests and other measures of performance that have been made during the course of training.

With the use of performance tests it is possible to measure how well the trainees can perform jobs requiring the basic skills taught in the schools. This information is not only of value to the people who administer the schools but also of value to the people who receive the school graduates and must assign them billets.

6. **They Can Indicate Progress Within the School Curriculum.** Since most Navy school curricula are divided into units or phases, it is necessary to determine when a man has successfully completed a given phase. For example, under the Craftsman Method of teaching it is desirable to move the men along as fast as they can go, but at the same time to be sure that they have mastered one phase before they start another. Performance tests lend themselves to this kind of evaluation and provide important information for school administrators and instructors.

7. **They Can Help You Spot Strengths and Weaknesses in Training.** It is often desirable to know whether a man or a group of men are strong in certain skills and weak in others. This problem arises especially in shipboard training programs. For example, men are often assigned those duties which they perform best with the result that they get no training in other important required duties. If you are a training officer in an operating unit, you will want to know that your training program is covering those areas of weakness which result from this specialization.

If you are an officer in charge of a Navy school, or the chief instructor, you'll want to know whether or not all of the phases of training are being equally well taught.

Performance testing, because it is a method of indicating how well a man can perform tasks requiring the basic skills of his rate, is a valuable tool for finding strengths and weaknesses. Either afloat or ashore, with this information you can take corrective action to bolster up a training program.

8. **They Can Pre-test Men Entering School.** In many of our Navy schools or other training programs, new men often start out already able to do some of the important tasks that are to be taught in the training program. Under these circumstances, it is desirable to save the Navy's time and to avoid dulling the interest of the men by permitting them to skip over the training on those tasks they can already perform. By the use of performance tests it is possible to measure the proficiency of these men before schooling begins and place them in the stage of training from which they can benefit most.

In general, performance tests can be used either to determine whether or not men meet the **minimum requirements** of performance on a given job, or to **discriminate** between men in their performance on that job. For example, if, in a radio school, a standard is set that requires all men to receive code at the rate of twenty words per minute, then, a good performance test can tell whether or not the students have met this minimum requirement. On the other hand, if you wish to assign scores to a group of radiomen to indicate their **relative** ability to receive code, then a performance test can tell you which of those men is able to receive code fastest or with the fewest errors, which the next, and so forth on down.

The actual use to which you put the test will depend on what you want to do with the results. If you want to know whether or not a man is sufficiently well qualified to be put up for advancement in rating, or if you want to know which man out of a group of men is most qualified to go up for advancement, then the performance test can be designed to fulfill either requirement.

CHAPTER II

Test Tasks and Types of Tests

INTRODUCTION

The tasks which are set up for use as performance tests will depend on what you want to do with the results. If you wish to know whether or not a man can stand a watch on a given piece of gear, it is only necessary to test him on that equipment. If you want to determine whether or not a man is qualified for advancement in rating, then the tasks should come from the practical factor requirements of the rate he is going up for. If you want to know which men are ready to begin a new phase in school, then the tasks for testing should come from the school curriculum. In short, the choice of test tasks will depend on what you need to know about the examinee's ability.

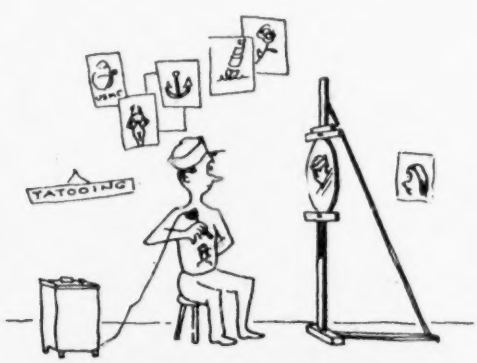
WHAT SHOULD THE EXAMINEE KNOW HOW TO DO?

The **first step** in building performance tests is to determine what the men you are going to test should know how to do. Depending upon your purposes in testing, you will want a list of rate, billet, or school requirements which will serve as a foundation for the **second step,** selecting and devising test tasks. If you only want to test the examinees on an existing routine task that they should be able to perform, after you have completed the first and second steps, all you have to do is build a test around that task. If, however, you are testing men for advancement in rating, or achievement in school, the requirements will include a **list of tasks** the men should be able to perform and a **list of skills** they should possess.

In this manual we will use the term "Task" to refer to the jobs that a man performs in his billet. The term "Skill" will be used to refer to the underlying basic abilities that enable a man to perform these tasks.

Here is a list of representative tasks and some of the skills which underlie them:

RATE	TASK	SKILLS INVOLVED
Machinery Repairman	Turning out a machine screw	Using a lathe; blueprint reading
Metalsmith	Making a locker	Welding; riveting; blueprint reading
Radioman	Receiving a message	Code receiving; typing
Cook and Baker	Baking a pie	Mixing dough; using an oven
Quartermaster	Getting a fix by bearings	Using pelorus; chart reading; plotting
Electrician's Mate	Repairing a lighting circuit	Circuit trouble shooting; using a megger; reading a wiring diagram
Yeoman	Preparing a letter	Typing; using standard procedure



TASKS REFLECT SKILLS

It is important that your list of requirements include **skills**, since a man with a good grasp of the basic skills of his rate will be able to apply them more efficiently to a wider variety of specific tasks. The list should also include the more important **tasks**, so that no area of performance will be neglected in making up the performance test battery.

Depending upon what you want to know about the examinees, you will need to consult one or more of the following sources of information to make up your list of tasks and skills:

1. **NAVPERS 18068, The Manual of Qualifications for Advancement in Rating.** This manual will give you a general listing of practical skills required for each rate and each pay grade in the Navy. It indicates what each man should know how to do to be qualified in the practical factors of a given rate. If you want to test men's qualifications for advancement in rating, then the "practical-factors" section of NAVPERS 18068 will be one of your principal sources of information. However, you should use other sources of information too.

2. **Experienced Personnel.** Chief petty and warrant officers can give you important information concerning the tasks and skills needed to perform the duties of a rate. Moreover, if you wish to determine whether or not one man can fill a given billet, or if you are interested in determining which man out of a group of men can **best** fill a billet, then men with experience in that billet are your best source of information. For example, if you wish to make up a battery of tests for a yeoman's billet, then your best source of information is the Chief Yeoman. You can also go to division officers to get additional task and skill requirements.

3. **Training Curricula.** To build performance tests for training school or on-the-job training evaluation, a list of tasks and skills can be obtained by looking over the training curricula. The trainees should be tested on the materials taught.

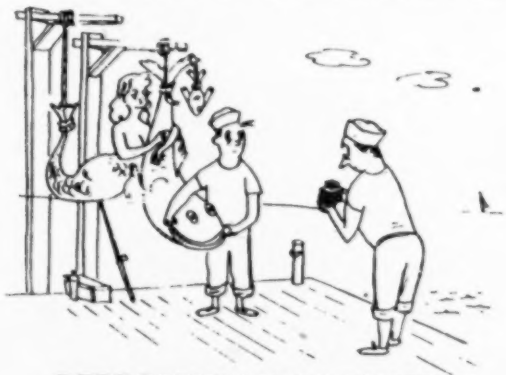
4. **Navy Training Courses.** The training course manuals published by the Bureau of Personnel provide information concerning the practical duties of the various ratings. The test builder without experience in the rating for which he is building tests should consult the **Navy Training Course** manual for that rating.

THE TEST TASKS

After you have made up the list of tasks and skills, you will need to decide which ones to use for tests. In some instances you will be able to use routine daily tasks **as they are** for the test tasks. More often, however, you will either have to **modify** the actual task to fit your testing needs, or you will have to **devise** a test task. This latter will be necessary if you wish to measure ability in a broad area (for example, "Generators") or a general skill (for example, "Welding"). Devising the test task is part of performance test construction and is covered in each of the three chapters that follow on how to build tests.

Whether you devise the test task, modify an existing task, or use the existing task as is for testing purposes, there are certain considerations to be kept in mind when selecting test tasks:

1. **Representative Sample.** If the list of tasks and skills is extensive, it is usually not feasible to test



REPRESENTATIVE SAMPLE

the examinees on all of the tasks. Therefore, the test tasks should cover a representative sample of the skill requirements. That is, there should be one test task covering each of the important skills, and there should be little overlap between the skills which are reflected by the tasks.

2. **Important Tasks.** The test task itself should be important. It, or tasks like it, should be frequently required in the performance of the billet or rate being tested.

3. **Practicality.** "Is the test task practical?" Is it economical? If the examinee is to turn out a product, can you get the materials needed for the work? Does the task take too much time? Is the needed equipment available? All of these practical considerations affect your choice of test tasks.

4. **Number of Men Required to Do the Test Task.** The examinee should be able to perform the test task without the help of other men. Tasks which require the work of more than one man for their completion should be avoided. The reason for this is that in these situations it is difficult to determine the adequacy of the examinee's performance independently of those helping him.

5. **Difficulty Level.** The test task should not be so easy that all of the examinees can perform it perfectly, yet it should not be so difficult that none of the examinees can perform it. The basic assumption of testing is that there are differences in men's ability to perform. The test task should be just difficult enough to separate the poorer from the better men.

Thus after you have made up your list of tasks and skills, you screen the list to eliminate those which cannot be made into good test tasks. (Of course not all of these standards can always be met. However, an effort should be made to have your final test tasks conform to as many of these standards as possible.) You are now faced with the problem of selecting the type of performance tests to build. The remainder of this chapter describes the various types of performance tests and their uses.

DECIDING WHAT TYPE OF TEST TO BUILD

There are three types of performance tests described in this manual. The type to use in any particular instance will depend on the test task, on the administrative limitation of your own situation, and on what you want to know about the examinees. Any of the three types can be built around some tasks, and the one to select is the one which will best tell you most about the examinees and still be administratively possible to use. In some cases two of the types can be combined and built into a single test of a specific skill or group of skills. In other cases one type of test will fit a particular test task and your needs best. Sometimes, because of administrative problems (for example, the number of men to be tested) you are limited to using a particular type of test. In this case, you must select or devise test tasks which fit that type of test.

The goal of performance testing is to translate a man's performance on a practical task into a score which reflects the quality of his performance. The **record** of performance is the means by which performance is translated into scores. This manual covers three basic ways of recording performance, and performance tests get their names from these three **recording methods**: Final Product, Check List, and Examinee Recorded.

The record of performance for each man tested serves several purposes besides providing means for assigning scores:

1. It provides the information that stands for the examinee's performance.
2. It permits each man to be scored on the same basis and thus helps minimize instructor or supervisor bias or prejudice.
3. It usually permits the examiners to score the tests at some convenient time after the test is completed.

An examination of each of the recording methods will show how each is used, where each is best used, and what some of the advantages and limitations of each type can be. With this information the test builder will know what type of test to use in any situation he might face.

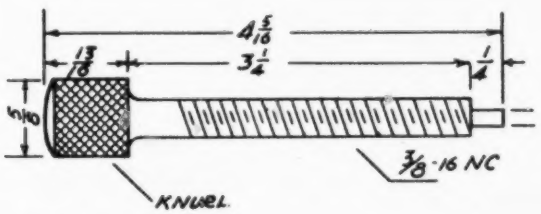
THE FINAL PRODUCT

The "Final Product" method of recording performance does not require an examiner to watch the

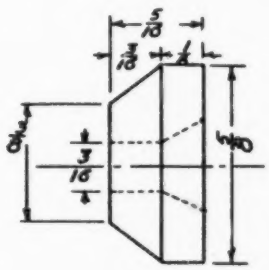
examinee perform each step in the test task. This method of recording applies to tests in which the examinee's performance results in something produced, or repaired. The result of the examinee's performance is some sort of a product, which, by itself, is a record of what he did and how well he performed. The following are examples of what is meant by final product performance tests:

1. Writing a letter
2. Soldering a wire splice
3. Welding a seam
4. Filling in a requisition
5. Repairing a circuit
6. Making a log entry
7. Building a tool box
8. Fabricating an aluminum/flush patch
9. Making a fender

If the skill around which you are building a test results in a final product such as those above, and if you are not interested in how the examinee cares for tools, the procedure the examinee follows in doing the job, or whether the examinee adheres to prescribed safety precautions, the best approach to recording performance is through the final product itself. This approach reduces the length of the formal recording form and also reduces the possibility of examiner bias. An instruction sheet is made which clearly indicates to the examinee just what kind of a product he is expected to turn out. For example, a final product test built around a man's skill at operating a lathe might utilize a drawing such as that shown below with instructions covering the tools and machines to be used.



BUTTON FOR "C" CLAMP



The examinee performs on the lathe, turning out the product as specified. The product that he turns out is a record of his performance, and it is evaluated in terms of the degree to which the specifications were met, together with as many other features of the work as properly reflect excellence of performance.

THE CHECK LIST

The "Check List" is a series of statements, each of which defines a step in the test task. An examiner watches the examinee perform the test task and records on the check list whether or not the examinee successfully completes each step.

The check list type performance test lends itself best to the type of tasks that require a man to go through a series of steps in the proper sequence for successful completion. Examples of this kind of task are as follows:

1. Starting an engine
2. Tuning a transmitter
3. Paralleling alternators
4. Putting a generator on the line
5. Operating a winch
6. Lighting off a radar

If diagnostic information (such as determining whether or not the examinee observes safety precautions and takes care of the equipment and tools, or determining where he is weak and falls down on the job) is wanted, then the check list performance test is the one to use. An example of part of a check list recording form for a Transmitter Tuning Test appears below. Notice that the examiner's job is to observe the examinee and simply circle "1" or "0" (or in some cases "Yes" or "No"), according to whether or not each step was successfully completed.

PART IV ANTENNA COUPLING: TUNE

1 0 13. Increase antenna coupling for 70 ma. of P.A. plate current.
1 0 14. Re-adjust P.A. tuning control J for minimum P.A. plate current.
1 0 15. Re-adjust antenna inductance for maximum P.A. plate current.
1 0 16. Re-adjust antenna capacitance for maximum P.A. plate current.

PART V ANTENNA COUPLING: OPERATE

1 0 17. Shift Adjust-Tune-Operate switch to operate position.
1 0 18. Re-adjust P.A. tuning control J for minimum P.A. plate current.
1 0 19. Re-adjust antenna inductance for maximum P.A. plate current.
1 0 20. Re-adjust antenna capacitance for maximum P.A. plate current.
1 0 21. Is P.A. plate tuning at a minimum with plate current at approximately 100 ma.?

The "check list" approach to recording performance can be used to make a record of performance on any job which can be broken down into a series of steps which must be completed successfully to do the job. Consequently, there are a great many jobs in the Navy around which check lists can be constructed and for which this approach is very convenient. However, there are some situations in which use of a check list should be avoided. Generally speaking, the main disadvantage to the check list approach is that it requires one examiner for each examinee. It also relies on the examiner's ability to judge correctly the performance of the examinee as it occurs. This is not always an easy job and introduces the possibility of examiner bias. Therefore, if the task lends itself to some other kind of recording, that is, either the examinee recorded approach or the final product approach, then the check list method of recording probably should be avoided. In some cases, the check list approach can be combined with the final product approach.

COMBINING THE CHECK LIST AND THE FINAL PRODUCT APPROACH

You may be interested in the manner in which an examinee does his job, as well as the quality of the final product he produces. In this case the check list approach to evaluation of performance in process may be combined with the final product approach of evaluating a man's end product. For instance, suppose you wanted to test a man on his ability to fabricate an aluminum flush patch. You might be interested in the procedure the man follows in doing the job, his method of handling tools, and the safety precautions he observes. At the same time, you might be interested in the adherence of the final patch to prescribed dimensions and its freedom from defects. In this case you could construct a recording form composed of statements referring to the examinee's procedure as well as to statements referring to the freedom from defect of the final product the man produces. Sample statements from a combined approach are shown below:

TOOLS

1. Always uses back-up plate when center punching 1
2. Does not elongate drill holes with drill 1
3. Checks dividers when finished with measurement transference to make sure dividers did not slip 1

Etc.

PROCEDURE

12. Lays out and cuts out damaged area 1
13. Rough files 1
14. Smooth files 1
15. Lays out rivet pattern 1

Etc.

SAFETY

19. Uses goggles when drilling 1
20. Does not aim compressed air hose at self or others 1
21. Places handle on file prior to using 1

FINAL PRODUCT

26. Clearance between insertion and skin as specified (tolerance ± 1/16") 3
27. All corner radii 1/4" 3
28. Not more than 1 rivet head marred 3

EXAMINEE RECORDED

In many instances it is possible to assign the examinee a test task to perform, and he can make a record of his own performance on a sheet of paper. These are called "Examinee Recorded" performance tests. The examinee recorded performance test frequently is the best choice in situations where the examinee is required to make measurements, to determine locations of casualties, to make interpretations of data, or to identify machinery and tools. The following examples may serve to illustrate these kinds of tasks.

1. **Measurements.** A quartermaster taking a bearing with a pelorus; an electrician's mate measuring the resistance to ground in a cable.
2. **Locations.** (Locating casualties in mechanical, electrical, or hydraulic systems.) An electronic technician locating a short in a radar set; an engineman locating a worn fuel injector.
3. **Interpretations.** An engineman interpreting the dial readings (on an instrument panel which might reflect normal or abnormal operating conditions); a metalsmith interpreting a blueprint.

Tasks of these kinds result in the examinee recording, during his work, some sort of information or an answer to some problem on a sheet of paper. He may record this information by making a check mark, encircling an alternative, or by writing in a numerical value or a descriptive word. The information recorded in this way serves as a record of the examinee's work and its correctness permits us to determine how well he did it.

The recording form used by an examinee when taking a practical performance test is seldom like the conventional answer sheet used with written tests. Usually it must be specifically designed to fit the particular test task at hand. For this reason examinee performance test recording forms usually differ widely from one test to another, and no standard form can be specified.

The examinee recorded performance test is a useful measuring device when a large number of men must be tested in a short period of time. Several examinees can be tested by one examiner. In addition, it is an objective measure of performance, since the examinee makes his own record of performance.

SCORING

There is another consideration which affects the selection of the type of test to be built around a task. This is the problem of scoring practical performance. Generally speaking, the ways in which you can record performance on a test task are more limiting, as far as test construction is concerned, than the ways in which you can score the performance. Therefore, the first question to answer in selecting the type of test to use is, "How can I get a record of the men's performance?" But, at the same time, consideration must be given to how that record of performance can be scored to reflect the men's actual performance.

It might appear that final product tests, because they automatically yield a record of performance, are the easiest kind of performance test to build. This is not always true. In many cases, it is more difficult to derive a meaningful score from a final product than it is from other performance records.

All three approaches to recording performance have advantages and disadvantages. Some of them have been pointed out in this chapter, but you may run into others in building tests. The particular test task you select, and the situation in which the test must be administered, will determine in large part which recording and scoring method you select.

SUMMARY

Various ways of determining what to use for test tasks have been presented in this chapter. In addition, a brief introduction to the types of performance tests was given. The chapters which follow deal in detail with the construction and use of these performance tests. Chapter III covers final product performance tests; Chapter IV deals with check list tests and combining final product tests with check list performance tests, while Chapter V describes the construction of examinee recorded tests.

CHAPTER III

Final Product Performance Tests

INTRODUCTION

Chapter II pointed out the difference between tests of **Final Products, Check List Tests,** and **Examinee Recorded Tests.** Here we consider final product tests only.

THE PRODUCT STANDS AS A RECORD OF PERFORMANCE

The general steps in building final product performance tests are as follows:

1. Determine the type of product the examinees should turn out to demonstrate the skills you wish to test. (In the case of a specific task, this step is omitted, since you already know the nature of the product the examinees should turn out.)
2. Devise methods of scoring the product.
3. Devise instructions for the examiner and the examinee.

DEVISING THE PRODUCT TO BE MADE

The first step in constructing final product performance tests is to decide what kind of product the examinees are to make, repair, or maintain, to demonstrate their skills. The products the examinees turn out should:

1. Require a broad sample of the skill you wish to test. The man who has a better grasp of a skill will be able to apply it to a greater variety of tasks. Consequently the test task should require as many aspects of the skill as possible.
2. Lend themselves to simple and objective methods of scoring.

OBTAINING A BROAD SAMPLE OF PERFORMANCE

If you can devise a test task that requires the **skill** necessary in performing a **number** of daily routine tasks, you will get a better indication of how well the man can fulfill the duties required by his billet.

You can get a broad sampling of the examinees' skill in two ways:

1. Specify one product for the examinees to turn out which will require several aspects of the skill being tested. For example, to test men's ability to use a lathe you might have them turn out a product whose construction requires turning, boring, facing, and screw cutting. In this way they would have to demonstrate a broad sample of their ability to use the lathe.
2. Specify **several** products for the examinees to make, **each** of the products requiring a **different aspect** of the skill being tested. For example, instead of having the machinery repairmen demonstrate their ability to use the lathe by turning out **one** product, have them turn out **four** products. One of the products could call for them to demonstrate ability to use the lathe for turning, another product call for boring, the third for facing, and so on.

Regardless of whether one or several different products are to be made by the examinees, increasing the number of product samples that the examinees make increases the **reliability** of test scores. If, for example, men are being tested on their ability to weld, and test scores are based on only **one** sample of each man's work, there is a chance that some of the poorer men will accidentally do well on the test and some of the better men may not perform up to their usual standards. If, however, scores are based on **several** samples of each man's work, the chances are better that the scores will reflect each man's ability properly. By increasing the number of samples on which you base test scores, the examinees gain an opportunity to settle down and do the quality of work they customarily do, thus increasing the degree to which your scores reflect the facts.

SCORING FINAL PRODUCTS

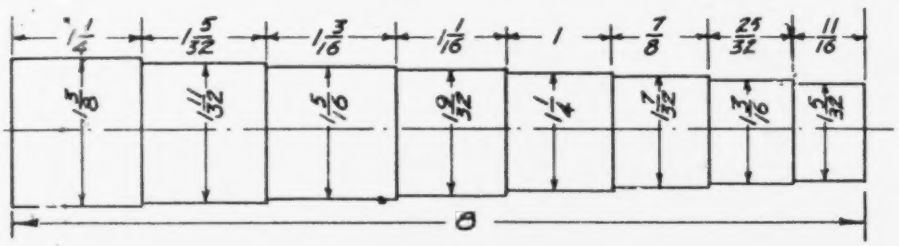
When specifying the product the examinees are to make to demonstrate their skill, consideration should be given to how that product can be scored. The product sometimes can be designed in such a way that it can be scored objectively and conveniently.

Different products must necessarily be scored in different ways.

There are five ways final products can be scored; some products can be scored by using only one of these methods and others by using several of them.

1. **Physical Measurement.** In many instances the examinee will be given instructions to turn out a product to specifications. That is, he will be told to make something to given physical dimensions with a certain degree of tolerance. The products can then be scored on the basis of whether or not they meet the specifications. One point is given for each dimension that is within the given tolerance limits.

For example, machinery repairmen might be required to turn out the shaft pictured below.

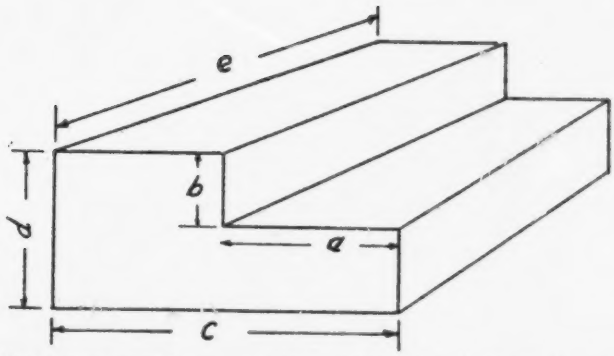


Their performance could be scored on sixteen different measurements: eight diameters and eight lengths. One point could be given for each measurement that is within the given tolerance limits.

It may be possible to score final products in terms of the **degree** to which their physical dimensions meet the specifications when "no tolerance" is specified. Suppose machinery repairmen were told to turn a shaft down to two inches diameter, no tolerance. Those diameters which are within one ten-thousandth of two inches might be assigned a score of three, those within one five-thousandth a score of two, those within one-thousandth a score of one, and those diameters outside one-thousandth a score of zero.

Two precautions should be pointed out here. First, if the man's products are to be scored in terms of the degree to which they meet the specifications (how close they are to perfect), the men should be told to make the product as close as they can to the specifications. They should be told that there are no tolerances. Second, precautions must be taken, when scoring products by measurements, to insure that the machines and tools the men use are equally precise. This can be done by having all the examinees use the **same** tools and machines or by having each examinee turn out one product to the same specifications **using each** of the machines and sets of tools. Each man's score for that test task would be obtained by averaging the scores of all his products. In other words, each man should have the opportunity to use both the good and poor equipment or else all men should use the same equipment.

Another way of deriving scores on the basis of measurements of physical dimensions is to obtain the **absolute sum** of the **errors** or **discrepancies** from the specifications. By **absolute** is meant the value of the discrepancy disregarding whether the measured dimension is too large or too small. An example will make this clearer. Assume you were testing men's ability to use a shaper and had them turn out the notched block pictured below:



Perhaps to score these blocks you would measure the width (a) and depth (b) of the notch and the width (c), thickness (d), and length (e) of the block, making five measurements. You could derive a score by adding together the amount of discrepancy, found in each of these measures from their prescribed values, disregarding the direction of the discrepancy.

2. **Errors or Freedom from Defects.** For certain types of products, physical measurements may be combined with freedom from defects as a means of scoring. For example, if the products are such things as products turned on a lathe, a chair constructed by a carpenter, or a patch in an aircraft's surface, their quality is determined in part by their freedom from defect. In these cases points can be subtracted from an examinee's score for each error that is made. In some cases, errors may be used as the primary means of scoring. For example, if the products are such things as typewritten letters or sent or received messages, their quality is determined in part by their freedom from errors. One point can be subtracted from a perfect score for each error that is made.

In many instances, physical dimensions alone do not give a complete picture of the quality of a product. A gear, for example, may fit the dimensional specifications, yet have pits, flaws, or tool marks on some of its surfaces. It would not compare in quality to another gear of the same dimensions without these flaws. Points should be deducted for non-measurable errors such as these. The important point to remember is to score all errors on the same basis, deducting the same number of points for errors of equal seriousness, and to be sure that examiners look for the same errors in each final product.

Sample items from an examiner's "reminder sheet" for scoring a plastic patch by aviation structural mechanics follow:

20. NO bubbles in joint	3
21. NO scratches or scars on plug	3
22. NO dirt in joint	3
23. NO spaces in joint	3

In some cases the final product will be such work as a sample entry made in a ship's log, a letter written by a yeoman, a stub requisition filled in by a storekeeper, or a sample service record entry. An aid that has been found valuable for scoring end products such as filled-in forms is shown in Exhibit I. This aid was constructed in order to help examiners score data sheets which were filled in by Naval photographers. A portion of the examiner's check list is shown below:

3. Location—"at sea" or fictitious location stated	1
4. Ship—Name and number stated	1
5. Date Shipped—Fictitious date stated; sequence of day, month, and year followed	1
6. Photographer—Name and rate stated	1

EXHIBIT I

Scoring Aid Used in Scoring Photographer's Data Sheets

PHOTOGRAPHER'S DATA SHEET

UNITED STATES NAVAL PHOTOGRAPHIC SERVICE

NAVAER-1086 (REV. 12-45)

Fill out in triplicate

LOG No. (DO NOT WRITE IN THIS SPACE. FOR FILE USE ONLY)

INDEX No.

SUBJECT (*Give project number, if any*)

Scenes of Crash and Rescue at Sea

LOCATION	SHIP OR STATION	DATE SHIPPED
At Sea	U.S.S. RANDOLPH (CV-15)	9 September 1952
PHOTOGRAPHER	**SIZE OF ROLLS, NEGATIVES**	**TOTAL THIS SHIPMENT**
A. B. SEE, AP1		4 Negatives

TYPE OF FILM (*Check one box only. Use separate data sheet for each type of film*)

☐ BW, 16 MM. ☐ COLOR, 16 MM. ☐ BW, 35 MM. ☐ COLOR, 35 MM. ☑ BW, STILLS ☐ COLOR, STILLS

CAMERA TYPE	CAMERA SERIAL NO.
K-20	18068

GENERAL INFORMATION (*In addition, state any unusual conditions or problems. If made with sound, state system—wire, film, disk, etc.*)

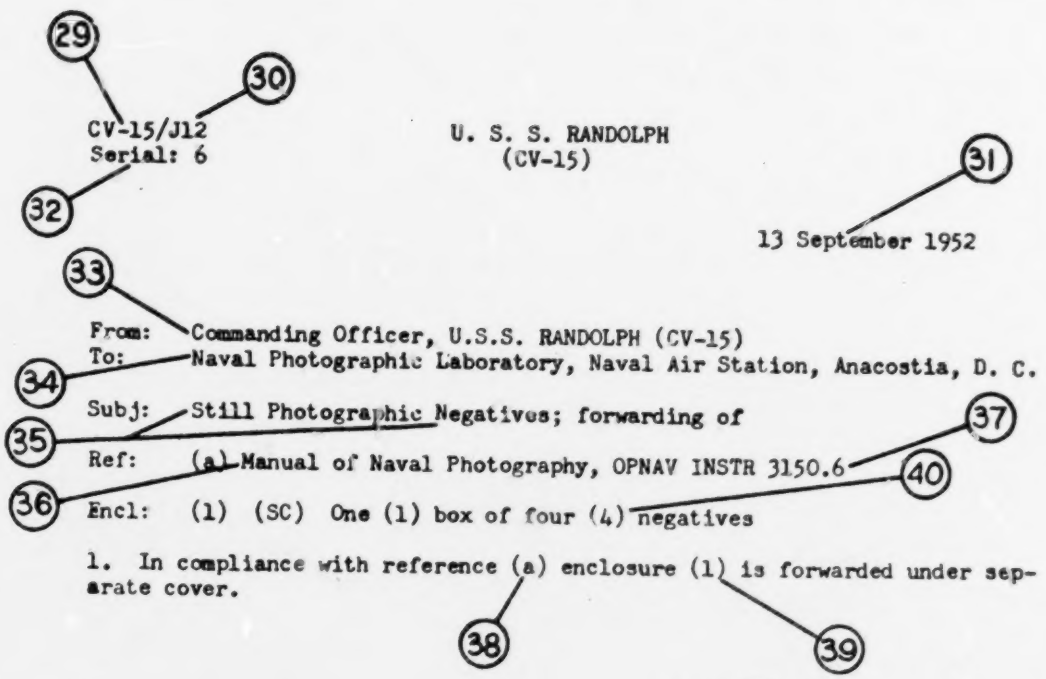
Crash occured because of engine failure on catapulting.

ROLL AND SCENE NO. OR PACK AND NEGATIVE NO.	SUBJECT DESCRIPTION (*List scenes, names, and events accurately. Attach newspaper clippings, programs, speech copies, Plan of the Day, etc.*)	DATE AND TIME OF IMPORTANT EVENTS
Negative Nos. CV-15 1	F4U crashed into sea off coast of Korea. Pilot LT J. DOE, USN climbing from cockpit.	0800, 9 Sept. 1952
2	Escort Destroyer, U.S.S. SMITH (DD-120) launching small boat to pick up pilot.	0804, 9 Sept. 1952
3	Pilot LT J. DOE, USN aboard carrier receiving medical aid from carrier medical officer, CDR D. MANY, USN	0815, 9 Sept. 1952
4	Aerial oblique: Port Side: U.S.S. RANDOLPH (CV-15) at sea; altitude 500 feet: Exposure data: 1/500 sec. @ f/8	9 Sept. 1952, 0900

16

17

EXHIBIT II
Scoring Aid Used in Scoring Letter Writing Performance Test



(29) (30)
CV-15/J12
Serial: 6

U. S. S. RANDOLPH
(CV-15)

(31)
13 September 1952

(32)

(33)
From: Commanding Officer, U.S.S. RANDOLPH (CV-15)
(34) To: Naval Photographic Laboratory, Naval Air Station, Anacostia, D. C.

(35) Subj: Still Photographic Negatives; forwarding of

(37)
Ref: (a) Manual of Naval Photography, OPNAV INSTR 3150.6

(36) (40)
Encl: (1) (SC) One (1) box of four (4) negatives

1. In compliance with reference (a) enclosure (1) is forwarded under separate cover.

(38) (39)

A. B. SEE

In order for the examiner to score a data sheet, the examiner simply checked for the presence of each of the bits of information called for in the examiner's check list. If the examiner was hazy as to what was meant by an item, say #4, all the examiner had to do was to look at #4 on the correctly filled in sheet (Exhibit I) and there he saw a correctly filled example. The same type of scoring aid was provided to help examiners in scoring a covering letter, prepared by examinees, for the data sheets. Part of the examiner's check list for the covering letter is shown below while the sample letter that was provided is presented in Exhibit II.

29. Correct number of initiating source stated 1
30. Correct or fictitious file number stated 1
31. Date stated in proper sequence 1
32. Correct or fictitious serial number stated 1

3. **Scale of Model Products.** There are certain types of final products which do not easily lend themselves to scoring on the basis of physical measurements. It would be practically impossible, for example, to determine the quality of a solder or a wire splice without making very elaborate measurements. The quality of a soldered wire connection could be determined by measuring the resistance it offers to the flow of electrical current. And, the quality of a wire splice could be determined by measuring its strength with a materials testing machine. But these methods are generally administratively impractical.

In a practical testing situation, products of this type can be scored by using a scale of **model products.** This scale consists of a series of actual products similar to the ones the examinees are required to make. To construct the scale, you rank the models according to their quality from the best to the poorest and assign them scores according to their rank. For example, to score solders, a scale of solders from excellent to poor could be set up like this:

MODEL SOLDERS

Sample	Points	Description
	4	Excellent
	3	Good
	2	Average
	1	Poor
	0	Unsatisfactory

Each examinee's test product is assigned the score of the product in the scale which it most closely resembles.

The products used in the model scale can be obtained by having one experienced man turn out several products varying in quality or by accumulating samples of work of men who display the skill in varying degree.

Ranking the products according to their quality in order to obtain a model products scale may require the judgment of several experienced men.

4. **Does It Work?** Many products cannot be scored by making measurements and looking for defects or by comparing them to a scale of model products; there are no graduated degrees of quality in final products of this sort. **They either work or they do not,** and it is upon this basis that they must be scored. For example, if a gunner's mate assembles a gun as a test job, the final product either checks out properly or it does not; and it is on this basis that the product must be scored. If an electrician's mate connects a motor to a controller, the motor either runs or it does not, and there are no other degrees of

quality which can be scored. Thus one point is assigned the product which works and zero to the product which does not work.

DOES IT WORK?

There are many disadvantages to a test in which all the examinees receive a score of either zero or one. From the test scores you can determine which men can do the test job (those receiving a score of one) and which men cannot (those receiving a score of zero). But, nothing can be determined concerning the differences in ability among the men who made a score of one or those who made a score of zero.

The problem of scoring products according to whether or not they work can be solved by increasing the number of products the examinees make, repair, or assemble to demonstrate one skill. For example, to test electrician's mates' ability to connect motors to controllers, they could be required to hook up the motors in a number of ways: as a differential compound motor, as a cumulative compound motor, etc. Each hook-up would then be scored according to whether or not the motor runs properly: one point for each correct hook-up.

Increasing the number of products the examinees are required to turn out for one test has two important advantages. First, where the products are scored either one or zero according to whether or not they work, increasing the number of products to be turned out increases the total possible score and makes possible greater discrimination.

Second, regardless of which method of scoring is used, increasing the number of products the examinees must turn out increases the reliability of the test. In the case of the electrician's controller-motor test, a man might get one hook-up correct by chance but it is very unlikely that he would get ten hook-ups correct by chance.

5. **Time.** The time taken to make the final product may be used as a partial measure of a man's ability. Time as a measure, however, does not give an indication of the quality of the product, and it should not be used as the sole measure of performance unless all the products are of essentially the same standard of quality. A man should not be given credit for rapid work if the work itself is not good.

In some cases time **must** be used as a partial measure of performance. For example, a poor typist may type a letter with no errors if he is given all the time he wants. In this case time should be used as a measure of the proficiency of the man, or a time limit should be set and scores should then be based on factors other than time.

Time to make, repair, or assemble a product can be used as a measure of a man's skill when:

a. It is important to turn the product out rapidly when **on the job.**

b. The product can be turned out perfectly by even poor operators when given all the time they want.

c. Two or more examinees turned out products of equal quality but took substantially different amounts of time to do so.

Test time limits are discussed more fully in Chapter VI.

SUMMARY OF SCORING PROCEDURES

Final products may be scored in five ways:

1. By measuring their physical dimensions
2. By deducting points for errors or defective work
3. By comparing them to a scale of model products
4. By determining whether or not they work
5. By measuring the amount of time taken to make them

An important point to remember is this: If men are compared in terms of the scores they make on a test, each man's score must be obtained in the same way using the same scoring methods.

INSTRUCTIONS FOR THE EXAMINEE

Every effort must be made to see that the products are turned out under identical conditions. This means that the examinees must know what is expected of them: what they are to do, what materials, tools, and equipment they have to do it with, and the amount of time they have. Examinees' instructions should, therefore, contain the following information:

1. A statement describing the purpose of the test. For example: "This is a test of your ability to operate a shaper."
2. A statement of specifications describing the product the examinees are to make. For example: "You are to connect this motor to the controller in front of you to form a differential long shunt compound motor." A mechanical drawing or schematic of the product the examinees are to make is often the clearest description, and should be used whenever possible (provided the drawing or schematic does not supply so much information that the test becomes too easy). If there are tolerances to be observed, these should be clearly indicated.
3. A list of the tools and machines the examinees may use in performing the test job. For example:

   "TOOLS AND MATERIALS

   A. Knee and column universal milling machine

   B. Swivel vise and wrench

   C. Milling machine arbor and spacers

   D. Cutter

   E. Rack gear blank

   F. Vernier gear tooth calipers

   G. Parallels

   H. Dial indicator."

4. The time limits, if any. For example: "You have forty (40) minutes to perform this job."
5. The statement, "Are there any questions?" The test administrator must be sure that all the examinees understand what they are to do and should encourage questions in order to be sure that each man fully understands the task before he begins.

The examinees' instructions should be read aloud by the test administrator and at the same time the examinees should have copies of their own to look at closely before and during the test. Sample examinee instructions for the task involving the filling in of a Photographer's Data Sheet and composing a rough covering letter are shown in Exhibit III.

INSTRUCTIONS FOR THE EXAMINER

The instructions for the examiner should contain all of the information which the examinees' instructions contain, and, in addition, they should contain the following:

1. A description of how the final product is to be scored; the specifications for the products; the tolerance limits, the number of points to be deducted for errors; and how time scores are to be used (if any).
2. A description of how the testing area should be set up (This last point is covered more fully in Chapter VI.)

Sample examiner instructions covering the filling in of the Photographer's Data Sheet and composing a rough covering letter are shown in Exhibit IV.

EXHIBIT III

PERFORMANCE EXAMINATION

RECORDS AND REPORTS

SHIPMENT OF MATERIAL TO FILM LIBRARY

EXAMINEE INSTRUCTION

TASK — This is a test of your ability to prepare the forms required to cover mailing of motion or still picture material to the official U. S. Navy Photographic film library at the Naval Photographic Center, Anacostia, D. C. The material that you are to mail is described below. Your task is to fill out the Photographer's Data Sheet and to compose a rough letter of transmittal, both of which are needed to implement this shipment.

MATERIAL TO MAIL

You are the photographer aboard the U.S.S. RANDOLPH (CV-15). You have shot in black and white the following sequence of photographs and are mailing the negatives to the film library:

Negative 1 shows an F4U which on catapulting has engine failure and has fallen into the sea. The pilot is shown climbing from the cockpit.

Negative 2 shows an escort destroyer, the U.S.S. SMITH (DD-120), launching a small boat to pick up the pilot.

Negative 3 shows the pilot back on the carrier after rescue, receiving medical treatment from the medical officer.

In addition you are including in the shipment a fourth negative.

Negative 4 shows an aerial oblique of your carrier.

Any information which you might need but which is not supplied is to be filled in fictitiously. For instance, since you do not know when the photos were made, you may fill in any date as the date of the event.

All of your shots have been made with a K-20 Camera.

SCORING—You will be scored on the **QUALITY** of the finished forms.

TEST MATERIAL—The following have been provided for your use:

1. Pencil.
2. Pen.
3. Carbon Paper.
4. "Photographer's Data Sheets" (NAVAER - 1086).
5. 8½" x 11⅓" paper.
6. Photography Technical Bulletin — Cumulative Edition.
7. Photo Lab Index.
8. Photography (NAVPERS 10371).
9. Photography (NAVPERS 10372).
10. Paper clips.

EXAMINEE'S NOTES

1. This is a timed test.
2. Use the white 8½" x 11⅓" paper for all work other than work which you would ordinarily do on the "Photographer's Data Sheets."
3. Be sure to supply all necessary data, using fictitious data where necessary.

TIME — Time starts as soon as you understand what you are to do and report to the examiner that you know what your task is and are ready to start.

Be sure you understand what you are to do before you start. Have you any questions?

EXHIBIT IV

PERFORMANCE EXAMINATION

RECORDS AND REPORTS

SHIPMENT OF MATERIAL TO FILM LIBRARY

EXAMINER INSTRUCTIONS

TASK — This is a test of the examinee's ability to prepare the forms necessary for shipping motion and still picture material to the Central files of official U. S. Naval Photography at the Naval Photographic Center, Anacostia, D. C.

GENERAL DIRECTIONS TO EXAMINER

1. Read very carefully and thoroughly the "Examinee Instructions." Be sure you understand the test and the method for administering it before attempting to give it to the examinee.
2. Assign the examinee to a desk in a quiet, ventilated room and provide him with the test material.
3. Have the examinee read the "Examinee Instructions." Ask him if he has any questions. If he has a question, reread to him that part of his instructions which answers his question. Make sure that the examinee understands what he is to do before he starts. Supply no accessory information. Remember, this is a test situation, not a training one.

SCORING — You should receive each of the following products from the examinee:

1. Photographer's Data Sheet.
2. Letter of Transmittal.

Each of these is to be scored separately on the appended scoring sheets. Also appended is a sample data sheet, correctly filled in, and a sample covering letter. REFER TO THESE SAMPLES WHEN SCORING. DO NOT ATTEMPT TO SCORE THIS TEST WITHOUT THE USE OF THESE SAMPLES. If you think that the examinee should receive credit on an item, do nothing to the number which appears after that item. If the examinee is not to receive credit on an item cross out or X over the number appearing after that item. On completion of the scoring, add up the non-crossed-out numbers and enter the total in the space provided.

TIME — Time is counted as soon as the examinee understands what he is to do and is told to start. For credit on the time item, the examinee must finish in 45 minutes or less. If he does not finish in 45 minutes or less, allow him to finish. He receives no credit on the time item, however, unless he completes the data sheets and covering letter in 45 minutes or less.

SUMMARY

A wide variety of the practical skills which Navy men must possess can be tested using final product performance tests. Final product tests are ones in which the examinee's performance results in a product which can be scored.

The steps in constructing final product tests are as follows:

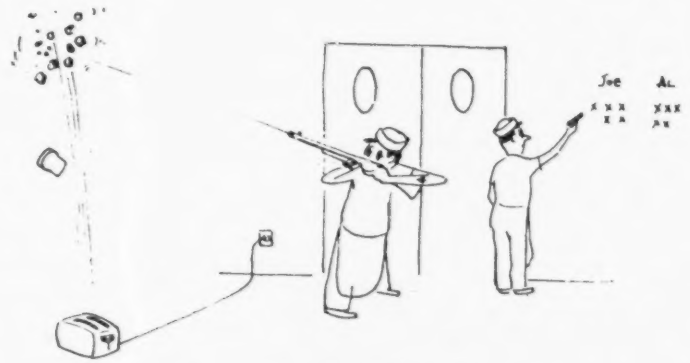
1. Determine the type of product the examinees should make, repair, assemble, or maintain to demonstrate the skill you wish to test. The product or products should require a broad sample of the skill you wish to test. This can be done by having the examinees turn out one product which requires many of the aspects of the skill, or by having them turn out several products each of which requires one aspect of the skill.
2. Determine how the product or products can be scored. There are five ways to score final products: by measuring physical dimensions, by comparing them to a scale of model products, by seeing if they work, by counting the number of errors, and by measuring the amount of time it takes to make them.
3. Write instructions for examinees and for examiners.

CHAPTER IV

Check List Performance Tests

INTRODUCTION

The check list performance test is used when you want to observe a man while he actually performs the test task. As a testing method it is simple and has a wide variety of uses. If you wish to test a man's ability to stand a watch on a given piece of equipment, you can use a check list and record each operation he performs. Or if you want to determine how well a man performs a task in which he makes something, you can watch him do the task, record and score his performance with a check list, and also score the final product he produces.



PERFORMANCE CHECK

Using the check list, not only can you score how well the examinee performs the test task, but you can also score how well he cares for tools and equipment and observes safety precautions while he performs.

With many tasks, the check list is the only method of testing that can be used. In these tasks no product is made or no information is gathered and hence the other types of tests covered in this manual are not applicable.

The tasks which are best suited to the check list generally involve what are called **procedural skills,** which are the sort of skills involved in performing the steps of a task in the **proper sequence** and usually in accordance with prescribed rules for safety and care of the equipment. A few examples of procedural type tasks are given below:

1. Starting an engine
2. Tuning a transmitter
3. Changing the plugs in an aircraft engine
4. Paralleling alternators
5. Putting a generator on the line
6. Loading an aerial camera
7. Field stripping a 20 mm. cannon
8. Operating the ship's boat

The only way that you can properly evaluate how well a man can perform tasks of this type is to watch him actually go through each step in the job in sequence. An examiner is required to watch the examinee perform the job and mark the check list according to whether or not each step is successfully and properly completed.

In many instances it is desirable to use the check list approach even though a final product results from the examinee's work. The check list is used in these cases because it reveals **more** about performance than an evaluation of the final product alone. For example, suppose the task were that of having a boatswain take the ship's boat to a dock, pick up some stores, and return to the ship. In a sense the final product of his performance is whether or not he got back to the ship with both the boat and stores in proper condition. So if one were to evaluate the final product of his performance that would be the only thing there would be to go on.

But there might be wide differences in the skill which various boatswains would show in accomplishing this same task. One boatswain might give poor orders to the engineman in getting away from the side of the ship, might have to reorient himself halfway out in the stream after having taken a series of courses which were not very appropriate, might violate several rules of the road enroute, might make a poor approach to the dock on landing, and make an equal number of errors on the return trip. Another boatswain might produce the same end result (same final product) but might execute all of the proceedings necessary for getting the boat to its destination and back again with a great deal of skill and dispatch.

The development of check lists for tasks from which no final product evolves is discussed in the first part of this chapter, while check list appraisal of procedure **plus** an evaluation of a man's final product is discussed in the latter part.

If you find, upon examining the list of skills and tasks around which you wish to construct performance tests, that one or more of them lends itself to the procedural check list approach to testing, the general test construction steps to be followed are these:

1. Devise a recording form which will contain all the **important** and **essential** steps required in performing the test task, the safety precautions the examinee should observe while performing, and the equipment care he should demonstrate while working.
2. Devise instructions for the examinee. These prescribe the task the examinee is to perform.
3. Devise instructions for the examiner.
4. Devise a method of assigning a score value to each step in the test task and to other features of performance such as safety precautions and care of tools and equipment.

DEVELOPING THE RECORDING FORM

The steps described in the sections which follow make up the "task analysis" which is performed in designing the recording form. Although scoring procedural check list tests is taken up in more detail later in this chapter, the features of performance that should be scored must be given consideration prior to the actual making up of the recording form.

Step 1. What features of performance will be scored: In general, the features of performance which can be scored when the check list test is used are as follows:

Completion of each step in the test task.

Care of equipment and tools.

Regard for safety rules.

Time.

The final product.

The recording form is made up with these scorable features in mind, so that items will be selected which can be scored in these terms.

Step 2. Rule a piece of paper as shown in Exhibit I. At the top of the paper write the name of the test task. Under this write as column headlines the features you are going to score. Your paper will then look similar to that shown below.

EXHIBIT V

TASK ANALYSIS — CHEMICAL MIXING

PROCEDURE	CARE & USE OF TOOLS	SAFETY PRECAUTIONS

Step 3. List the steps in the test task. If you are thoroughly familiar with the test task, and can correctly perform it yourself, list the steps to be performed in detail. If you are not completely familiar with the task, there are other ways you can get the steps.

a. List the steps an experienced and highly competent man goes through in performing the operation. This can be done by having such a person perform the task while you watch. Whenever questions arise he can explain to you exactly what he is doing and why.

b. Consult manufacturer's instruction book for the equipment concerned. One of the chapters in such a book is entitled "Operation" and this will indicate how, according to the manufacturer, the equipment should be operated.

In practice it would be desirable for the test constructor to follow all of these methods in determining the steps in the job. In this way he can be assured that he has not omitted any essential step and that he has not included any step that is not important to completion of the task.

Write in the steps of the test task under the "Procedure" column. List the tools and equipment that are used under "Care and Use of Tools." When you have done this your "task analysis" sheet will look like the one in Exhibit VI. This form was used in the development of a test of chemical mixing for aerial photographers.

EXHIBIT VI

TASK ANALYSIS — CHEMICAL MIXING

PROCEDURE	CARE & USE OF TOOLS	SAFETY PRECAUTIONS
Draw appropriate chemicals.	Dichromate, acid.	
Weigh out 3 ounce dichromate.	Scale, paper.	
Wash equipment.		
Add dichromate to small quantity of water.		
Stir while adding.		
Completely dissolve dichromate.	34 ounce graduate, stirring rod, tank.	
Add water to make one quart.		
Measure 12 ounces acid.		
Add acid to dichromate solution		
Fill quart container.	Quart container.	
Label (Name of solution, who mixed, date mixed).	Labels.	
Wipe down pans and work area.	Sponge.	
Replace weights and materials.		
Wash tank, graduate, etc.		

Next read over your list of tools and materials; ask yourself what the proper method of using each tool and material you listed is. Write in this information. In the case of chemical mixing your work sheet would now look like Exhibit VII.

EXHIBIT VII

TASK ANALYSIS — CHEMICAL MIXING

PROCEDURE	CARE & USE OF TOOLS	SAFETY PRECAUTIONS
Draw appropriate chemicals.	Dichromate, acid.	
Weigh out 3 ounces dichromate.	Scale — Weights on right, chemicals on left.	
	Check for "trueness" prior to use. Center counterweights.	
	Evenly distribute chemicals over pan.	
	Paper—Use clean paper on both pans.	
Wash equipment.	Rinse all equipment prior to use.	
Add dichromate to small quantity of water.	34 ounce graduate, stirring rod.	
Stir while adding.	Tank.	
Completely dissolve dichromate.		
Add water to make one quart.	Quart container.	
Measure 12 ounces acid.		
Add acid to dichromate solution		
Fill quart container.		
Label (Name of solution, who mixed, date mixed).	Labels.	
Wipe down pans and work area.	Sponge.	
Replace weights and materials.	Weights placed in proper places.	
Wash tank, graduate, etc.		

Step 4. List the safety precautions involved in doing the job. Having listed the procedure and tools and materials, you look over the lists and determine what safety precautions are involved. The question to ask as you read down each step of the procedure (and at the same time over the tools) is: "What can be done at this point that is unsafe?" List the safety precautions as they appear.

For the chemical mixing test, the appearance of the work sheet now would look like this:

EXHIBIT VIII

TASK ANALYSIS — CHEMICAL MIXING

PROCEDURE	CARE & USE OF TOOLS	SAFETY PRECAUTIONS
Draw appropriate chemicals.	Dichromate, acid.	Never leave uncapped.
		Resto immediately after use.
		Never pour acid in drain.
Weigh out 3 ounces dichromate.	Scale — Weights on right, chemicals on left.	

	Check for "trueness" prior to use. Center counterweights. Evenly distribute chemicals over pan.	
	Paper—Use clean paper on both pans.	Discard paper immediately after use.
Wash equipment.	Rinse all equipment prior to use.	Use only cold water.
Add dichromate to small quantity of water.	34 ounce graduate.	
Stir while adding.	Stirring rod.	
Completely dissolve dichromate.	Tank.	
Add water to make 1 quart.	Quart container.	
Measure 12 ounces acid.		Add acid at side of dichromate solution.
Add acid to dichromate solution.		Stir rapidly when adding acid.
		Add acid to dichromate, not reverse.
Fill quart container.		
Labels (Name of solution, who mixed, date mixed).	Labels.	
Wipe down pans and work area.	Sponge.	
Replace weights and materials.	Weights placed in proper places.	
Wash tank, graduate, etc.		

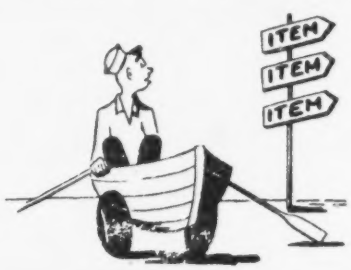
Sometimes there is more than one way of doing a task. If this is the case, it is necessary to list the two (or more) alternatives. For instance, when grinding a drill, some people completely grind one flute and then start on the other. On the other hand, some people alternate grinding from one flute to the other until they are completely satisfied with the job. Both procedures are acceptable and both would be listed in the task analysis.

REVIEW BY OTHERS

Once the task analysis is complete, have it reviewed by others who are proficient in the task analyzed and ask them to review it for accuracy. Ask the reviewers: "Is this how you would do this job?" "What other safety precautions are there?" "What other care of equipment characteristics does the good, careful worker show?", etc.

ITEM WRITING

The final recording form is constructed directly from the task analysis sheet and upon the recording form should appear some statement relating to each duty listed in the task analysis. The final recording form consists of a series of statements (items). An item is essentially a clear, simple, guidepost statement telling the examiner what to look for and score while giving the test.



GOOD AND POOR ITEMS

An item is a statement which describes a step the examinee should perform when doing the test or what the examiner should look for while recording the examinee's performance. As with everything else, there are good and poor items. In general, you can tell if you have a good item if you ask yourself these questions as you write.

1. **Is the item specific?** Does it refer to observable aspects of performance? Does it tell the examiner exactly what to look for? In the case of "precision of the final product," does it state what tolerance to allow?

Examples:

Good item — keeps clothing buttoned while working on plane.

Poor item — Clothing worn in safe manner.

Good item — Pats dry. Does not rub.

Poor item — Dries area or dries in proper manner.

Good item — Clearance as specified on schematic (tolerance $\pm$ $\frac{1}{64}$").

Poor item — Clearance as specified on schematic.

2. **Is the item worded simply?** Does it contain unclear, unfamiliar, or difficult terms? What else could the item mean to the examiner? Could unintended emphasis on a word or a phrase change the meaning? Is the item in any way apt to be misunderstood?

Examples:

Good item — Never touches lens with fingers.

Poor item — Refrains from digital or continuous contact with lens.

3. **Is the item in the correct working order?** Does it appear in the same sequence in which it appeared in the task analysis? In the case of alternative working procedures and tools, are all alternatives included on the final recording form?

4. **Are all scorable aspects scored?** What have I omitted? What else do people usually do wrong when doing this job? What other safety precautions are there, and which ones are usually ignored?

THE FINAL RECORDING FORM

On the basis of the task analysis for chemical mixing, the record sheet shown in Exhibit IX was prepared.

EXHIBIT IX

AF PERFORMANCE EXAMINATION
CHEMICAL MIXING
SCORING CHECK LIST

NAME_____ RATE_____

SERVICE NUMBER_____ DATE_____

Time Started_____

A. PROCEDURE

1. Draws correct chemicals 1
2. Weighs out exactly 3 ounces of potassium dichromate 1
3. Slowly adds dichromate to small quantity of water 1
4. Stirs while adding chemical 1
5. When all of chemical completely dissolved, adds water to make one quart 1
6. Measures out exactly 12 ounces sulfuric acid 1
7. Stirs rapidly when adding acid 1
8. Fills quart container with solution 1
9. Correctly labels (Name of solution, who mixed, date mixed) 1
10. Time finished_____. Time spent at mixing_____. Finished mixing in 8 minutes or less 1
11. Wipes scale pans and work area with damp cloth before securing 1

B. CARE AND USE OF TOOLS

12. Checks scale for "trueness" prior to use 1
13. Centers counterweights 1
14. Weights placed on right side of scale, chemicals on left 1
15. Wipes pans prior to use 1
16. Uses clean paper on both scale pans 1
17. Evenly distributes chemicals over pan 1
18. Rinses all equipment prior to use 1

C. SAFETY PRECAUTIONS

19. Never leaves acid bottle uncapped 1
20. Restows acid immediately after use 1
21. Does not pour acid in drain 1
22. Discards paper immediately after use 1
23. Uses only cold water in mixture 1
24. Adds acid at side of dichromate solution 1
25. Stirs rapidly when adding acid 1
26. Adds acid to dichromate solution (not reverse) 1

Total Score_____

SCORING CHECK LIST PERFORMANCE TESTS

The recording and scoring of the examinee's **performance** on a check list test task takes place as the examination is going on, while the scoring of any **end product** an examinee makes takes place after the examinee has completed his work. As seen in our chemical mixing example, performance check lists can be scored in several areas.

Completion of the steps in the test task. The total possible score on a check list test where there is no final product evolved is usually the total number of steps in the test task that are ordinarily performed. An examinee's score is then the number of steps he successfully completes **without help** by means of prompting or cues.

Sometimes more than one point may be given for a particular step. In such cases, of course, the total possible score could exceed the number of steps in the task. Usually it is not advisable to assign values other than **zero and one** to the various steps in a task, unless there is considerable evidence that certain steps are really more crucial in the total procedure than others.

Weighing the Steps. In the check list test devised around the task of tuning a transmitter, a form like that shown in Exhibit X shown below was devised. The scoring procedure for this check list originally consisted simply of encirling **zero** if the man failed to complete a given step without prompting, and encircling **one** if he did complete the step without help of any kind. Later the question arose as to whether or not some of the steps were not much more crucial than others and should receive special weights.

A number of highly experienced radiomen were interviewed, therefore, and it was determined that certain **phases** of the total operation were considered much more critical than others. Therefore, different **weights** were assigned for each of the phases in accordance with the average judgment of the experienced radiomen interviewed. The weights and the phases to which they were assigned were as follows:

Phase 1: **Preliminary settings:** 1½ points for each control on which the examinee made the correct preliminary setting.

Phase 2: **Master oscillator tuning:** 6 points for each step correctly executed by the examinee in this phase.

Phase 3: **Resonating all stages:** 3 points for each step successfully completed in this phase.

Phase 4: **Antenna coupling** (Tune): 5 points for each step successfully completed here.

Phase 5: **Antenna coupling** (Operate): 2½ points for each step successfully executed in this phase.

Part 6: **Final check by the examiner:** This was not a phase in the same that the others were but rather it was a final check to see whether or not the examinee had observed all of the critical features of the performance that he should. **One** additional point was added here for each question to which the examiner could answer "yes."

EXHIBIT X

NAME_____

RATE_____ CLASS_____

EXAMINER_____

RECORD SHEET
TRANSMITTER TUNING TDE (HF)
DESIRED FREQUENCY_____KCS
START
PART I PRELIMINARY SETTINGS

1. Set the following controls and switches in the indicated positions:

		CONTROL	POSITION
1½	0	HF-IF switch, control A	HF

EXHIBIT X — Continued

1½	0	Remote-Local switch	Local
1½	0	Adjust-Tune-Operate	Adjust
1½	0	CW-MCW-Voice switch	CW
1½	0	Antenna coupling control K	15-20
1½	0	Antenna tuning capacitor control M	50

2. Set the following controls to the position of the approximate frequency to be set up:

		CONTROL
1½	0	B
1½	0	C
1½	0	D
1½	0	E
1½	0	F
1½	0	G
1½	0	H
1½	0	J
1½	0	N

PART II MASTER OSCILLATOR TUNING

6	0	3. Press the start button and adjust filament voltage 10 volts.
6	0	4. Adjust plate voltage to 2000 volts.
6	0	5. Press test key on handrail and tune master oscillator, control C for zero beat in headphones.

PART III RESONATING ALL STAGES

(Test key must be pressed to tune all following stages)

3	0	6. Shift Adjust-Tune-Operator switch to tune position.
3	0	7. Adjust control E for maximum I.A. grid current.
3	0	8. Adjust I.A. tuning control G for maximum P.A. grid current.
3	0	9. Adjust P.A. tuning control J for maximum P.A. plate current.
3	0	10. Adjust antenna inductance for maximum P.A. plate current. (If no peak is obtained, does he shift ant. feed sw. to other position?)
3	0	11. Adjust antenna capacitance for maximum P.A. plate current.
3	0	12. Re-adjust P.A. tuning control J for minimum P.A. plate current.

PART IV ANTENNA COUPLING: TUNE

5	0	13. Increase antenna coupling for 70 ma. of P.A. plate current.
5	0	14. Re-adjust P.A. tuning control J for minimum P.A. plate current.
5	0	15. Re-adjust antenna inductance for maximum P.A. plate current.
5	0	16. Re-adjust antenna capacitance for maximum P.A. plate current.

PART V ANTENNA COUPLING: OPERATE

2½	0	17. Shift Adjust-Tune-Operate switch to operate position.
2½	0	18. Re-adjust P.A. tuning control J for minimum P.A. plate current.
2½	0	19. Re-adjust antenna inductance for maximum P.A. plate current.
2½	0	20. Re-adjust antenna capacitance for maximum P.A. plate current.
2½	0	21. Is P.A. plate tuning at a minimum with plate current at approximately 100 ma.?

STOP

PART VI FINAL CHECK BY EXAMINER

Directions: The examiner is to check the transmitter itself and answer each of the following items. Encircle yes or no.

YES	NO	1. Are filament and plate voltages correct?
YES	NO	2. Are any of the current meter readings too high?
YES	NO	3. Is each stage tuned to the correct frequency?
YES	NO	4. Are grid current meters peaked exactly?
YES	NO	5. Is P.A. plate current at exact dip?

It is difficult to say whether or not weighting the various steps in a task by different amounts, or weighting different phases by different amounts as was done in this example, will result in a better performance check list test than if simple weights of zero and one are assigned to each step. Frequently there is little difference in the standing a group of men will achieve on a performance check list with weighted steps as compared to one with simple unweighted steps. However, if the feeling of experienced officers and petty officers is that the different degrees of importance attached to the various steps be taken into account in the scoring, then the simplest method of procedure is as follows:

a. Pick out the **least important** step or phase that must be completed and assign it a weight of **one.** Assign a similar weight to all other steps which are roughly of the same difficulty level as this step.

b. Pick out the **most important** step or phase in the entire procedure. Obtain the opinions of experienced operators as to just how much more important this step is than the easy ones picked out first. It may turn out for example that performance in the most important phase is regarded as roughly four times as important as that in the least important phase. Therefore all steps or phases which are of the most important variety should receive weights of **four** as contrasted with those of **one** assigned to the least important steps.

c. Appropriate weights should now be assigned to all remaining steps that fall somewhere in between the important ones and the relatively unimportant ones. In this example all remaining steps should receive weights of either two or three.

You will probably find that weighting the steps in a procedure will help increase the discrimination ability of the check list test. That is to say, if only weights of zero and one are used, you may find that a number of men ended up with the same score. If weighted values are assigned to each step you frequently will find that individuals who formerly received the same score, or nearly the same score, may now have scores that distinguish quite a bit among them. Of course this makes it essential that proper weights be selected. We want to be sure that if two men have different final scores that the man with the greater score is really the better performer of the two.

2. **Care of equipment and safety precautions.** Since an examiner closely watches the examinee perform the test task, items covering the examinee's use of tools and equipment and his adherence to rules of safety are included in the check list.

By care of equipment we mean that the examinee properly cares for and maintains the tools used on a job, that the correct tools are used for the job, and that the tools are used correctly.

The importance of following the prescribed precautions for safety need not be elaborated. Failure to adhere to these regulations may result in the loss of time, bodily injury, and destruction of material.

3. **Time.** Another feature of performance that sometimes contributes to the scoring of a check list is the amount of time taken to complete the task. Thus, if a premium is placed on time in the **actual operational situation,** the examinee should receive a certain bonus for doing the task correctly within prescribed limits. This man's performance score should in some way differ from that of a man who does the same job but who takes much more time.

4. **Final product.** If the examinee's work on a check list performance test task results in a final product, it **also** may be scored. Generally speaking, a man who scores high on the check list should also score high on the final product, and vice versa. Consequently, the reasons for scoring both the procedure and the product then are perhaps not too obvious:

a. Scoring the final product alone does not enable you to get scores for "Safety" and "Care of Equipment."

b. The final product score does not tell you all of the things the man did while performing. The example of the boatswain at the beginning of the chapter brought out this point.

c. Scoring both the final product and the procedure permits the test to discriminate better. Two men might make the same score on the procedure section and yet turn out final products of sufficient difference in quality to enable you to give different scores to the two of them. The man who turned out the better product should naturally receive a higher score. Methods of scoring final products were explained in Chapter III.

Any of the methods of scoring final products can be combined with the check list scoring of performance, but probably the best combination is check list measurement of procedure plus final product scoring for freedom from defect and adherence of physical measurements of the final product to prescribed dimensions.

If your scoring includes measurements to be made on a final product then a fourth column is included on the "task analysis" sheet for these measurements. A complete task analysis for an aluminum butt weld, including final product measurements, is shown in Exhibit XI while the check list derived from the task analysis is shown in Exhibit XII.

Usually measurements of final products are made after the examinee has finished his work. However, sometimes defects in workmanship cannot be seen once the work is finished. For instance, if you asked a boatswain to prepare for paint and paint a surface, defects in the preparing of the surface might be hidden once the paint was applied. In cases like this it is well to make the appropriate measurements before they are concealed by the examinees.

EXHIBIT XI

TASK ANALYSIS — ALUMINUM BUTT WELD

PROCEDURE	CARE & USE OF TOOLS	SAFETY PRECAUTIONS	PRECISION OF FINAL PRODUCT
Select rods.	Rods — proper size for given metal thickness.		Start uniform with rest of weld.
Examine metal for dirt or grease. Clean metal and rods.	Metal—restricts cleaning to area of weld.		End uniform with rest of weld.
Set metal on jigs. Mix flux.	Flux — paste-like consistency. Free flowing.		Uniform penetration.
Flux base metal and rods.			
Adjust oxyacetylene regulators for 3-8 pounds.	Oxyacetylene bottle — 3-8 pounds.	Need fire extinguisher in area. Open oxyacetylene cylinders no more than 1½ turns.	Bead width 3 to 5 times metal thickness.
Light torch. Adjust to slightly carborizing flame.	Friction lighter. Torch—proper size tip. Goggles.	Don't use match. Hold torch on bench — don't aim in air.	Bead height 25 to 50% of metal thickness.
Pre-heat base metal.	Metal.		
Tack metal.			
Weld from center to end. Reverse metal and weld from center to other end using "see-saw" motion of torch and rod. Concentrate flame on base metal.		Loose clothing may be dangerous.	No irregularity in bead.
Dip and wash.	Water. Wirebrush.		

EXHIBIT XII

SAMPLE RECORD FORM FOR ALUMINUM BUTT WELD

AM PERFORMANCE EXAMINATION

WELDING

SCORING CHECK LIST

Time Started_____

TOOLS AND MATERIALS

1. Holds torch at 45° angle to work except for start or finish 1
2. Always concentrates flame on base metal, not on rod 1
3. Uses proper flux consistency (free flowing flux) 1
4. Selects proper size rod for given metal thickness 1
5. Selects proper size welding tip for given metal thickness 1
6. Restricts cleaning of base metal to width of weld 1

PROCEDURE

7. Examines metal for dirt or grease. Cleans both metal and rods 1
8. Sets metal on jigs. Mixes flux. Fluxes both base metal and rods 1
9. Adjusts oxygen acetylene regulators to 3-8 pounds (no credit if pressure on oxygen does not equal pressure on acetylene) 1
10. Lights torch and adjusts to slightly carborizing flame (feather should be no more than 1½ times inner cone) 1
11. Pre-heats base metal 1
12. Tacks metal from center to each end, or from center to each end alternately; tacks 1¼ to 1½ inches apart 1
13. Welds from center to one end 1
14. Reverses metal and welds from center to other end 1
15. Uses correct torch and rod motions while welding 1
16. Dips and washes making sure that all flux is removed 1
17. Time finished_____. Finished in 17 minutes or less 1

SAFETY

18. Shirt neck and sleeves buttoned 1
19. Makes sure fire extinguisher in area before igniting torch 1
20. Makes sure that gas bottles are in an upright position 1
21. Uses friction lighter to ignite torch and holds lighter on bench when igniting torch ... 1
22. Does not open acetylene cylinder valve more than 1½ turns 1
23. Uses goggles when welding 1

MEASUREMENT OF THE FINAL PRODUCT

24. Start of weld uniform with rest of weld 3
25. End of weld uniform with rest of weld 3
26. Uniform penetration for entire first 3 inches 3
27. Uniform penetration for entire last 3 inches 3
28. Bead width 3-5 times metal thickness for entire first 3 inches 3
29. Bead width 3-5 times metal thickness for entire last 3 inches 3
30. Bead height 25-50% of thickness for entire first 3 inches 3
31. Bead height 25-50% of thickness for entire last 3 inches 3
32. No bead irregularity in entire first 3 inches 3
33. No bead irregularity in entire last 3 inches 3

Total Score_____

EXAMINEE INSTRUCTIONS

An important part of your job in constructing a check list performance test is to devise an instruction sheet for the examinee which will specify very clearly for him the following information:

1. A description of the job he is to perform and just what he is required to do.
2. Information telling him what features of his performance will be scored.
3. A list of the tools and equipment he is permitted or expected to use.
4. The time limit, if any, including whether or not speed of performance is to constitute a part of the evaluation.

This instruction sheet should be carefully worded and the examinee should have a copy to read for himself. The examiner should also read the instructions aloud to the examinee before testing begins. After the instructions have been read, and after the examinee has had a chance to look over the details, he should be given an opportunity to raise questions about the task which may not be clear to him.

EXAMINER INSTRUCTIONS

Because of the relatively greater role that the examiner plays in check list performance tests, instructions for the examiner should usually be more elaborate than in the case of examinee recorded tests or final product tests. The examiner's instructions should contain the following information:

1. A brief description of the job and a statement covering precisely what is required of the examinee.
2. A list of tools and equipment the examiner must provide for the test and a description of any special features of the testing room or environment.
3. Instructions on how to observe the examinees perform, with special precautions for any steps which might prove difficult to observe.
4. Instructions indicating to the examiner what he should do in case the examinee:
    a. Starts to do a step incorrectly which could result in danger or which would throw all subsequent steps off.
    b. Forgets a step which must be done before he can continue.
    c. Does not observe a safety precaution which might result in injury to himself or the equipment.
5. Instructions on how to record performance by marking the check list appropriately.
6. The time limits if there are to be any.

Examples of examinee and examiner instruction sheets drawn up for the Aviation Metalsmith check list test on welding are shown in Exhibits XIII and XIV.

PROBLEMS IN USING CHECK LIST PERFORMANCE TESTS

Subjectivity. Since the score an examinee obtains on a check list scored only on procedure depends on an individual examiner and since there is no record of the examinee's work, there may be a certain amount of "subjectivity" in the performance check list. In addition, the more the examiner has to interpret what the examinee is doing, the more risk there is that his interpretation may distort the true facts of performance.

Nevertheless, experienced observers should be able to do a good job, particularly if the check list is constructed to call careful attention to all the features of performance that should be noted. Furthermore, we are faced with the fact that certain kinds of tasks lend themselves to this testing approach alone.

Hazardous Steps. Where possible it is best to let the examinee **do** the step incorrectly, checking him wrong on the recording form and letting him go ahead. There are steps in some tasks which when incorrectly done may result in damage to the equipment, injury to the examinee, or both. If the examinee begins to do something which is considered hazardous, he must be stopped, prompted on how to com-

plete the step correctly, and then permitted to go on. Of course he should not be given credit for that particular step.

Related Steps. There are steps in some tasks which must be successfully completed before a subsequent step can be undertaken or successfully completed. These are referred to as **related steps.** For example, if an examinee forgets step four or performs it incorrectly, he sometimes cannot do steps five, six, and seven. In this case the examiner should prompt the examinee on step four and not allow him credit for that step. He should then be permitted to go on and finish the operation.

If steps are not related and if no hazard is involved, the examiner should let the examinee go through the entire proceeding without prompting or stopping him. For each step that the examinee forgets or fails to do correctly the examiner simply encircles the "no" and withholds credit for the step.

Alternate Methods of Doing a Task. Sometimes in constructing a procedural check list test you will find that there is more than one procedure which may be followed with the same end result. In such a case it should be determined whether or not one alternative procedure should be considered the best one. If so, the check list should be built around this procedure. However, if it makes absolutely no difference which procedure is followed, but rather only that certain steps are completed, then an allowance for the various alternative procedures must be made in the check list.

Non-Essential Steps. Another problem which confronts the performance check list test maker is that it is sometimes difficult to determine which steps in the procedure are essential and contribute to the quality of the performance and which do not. The non-essential steps should not be a part of the check list. One way to determine which steps are, and which are not, really essential is to get the consensus of opinion of experienced officers and petty officers who know the job thoroughly.

Examiner Time. One final aspect peculiar to check list performance tests should be mentioned. Usually performance must be watched so carefully by the examiner that he can examine only one man at a time. This makes it difficult to use these tests in Navy schools, for example, where the desire may be to test a group of men at one time.

There may be real difficulties in arranging for enough time and examiners to get all men tested using a check list test. However, on ships and small installations it may be quite practical to use this method. Even in schools, procedural check list performance tests must be used because they are the only kind which lend themselves to certain practical tasks. In these cases about the only solution is to keep the total time required for the test to a minimum so that as many men as possible can be tested within the available time.

EXHIBIT XIII
AM PERFORMANCE EXAMINATION
WELDING
ALUMINUM BUTT WELD
EXAMINEE INSTRUCTIONS

TASK — This is a test of your ability to weld aluminum. You will be provided with two pieces of aluminum. **Your task is to butt weld the two pieces together along the length.** Do the weld as you would ordinarily do it. Be sure you understand what you are to do before you start. Make sure that you make the weld along the 6" side of the metal.

SCORING — You will be marked in the following areas:

1. The QUALITY of the finished product.
2. The PROCEDURE you follow in doing the job.
3. The MANNER in which you use your tools.
4. The SAFETY precautions you observe.
5. The TIME it takes you to complete the job.

The principal emphasis however, is on the quality of the finished product, so do not sacrifice doing a good job for speed.

TOOLS — The following have been provided:

1. Welding Bench
2. Oxyacetylene bottles
3. Welding torch and tips
4. Pliers
5. Flux
6. Steel wool, wire brush or emery cloth
7. Stiff bristle brush

If you need anything else, you may draw it.

MATERIAL — Two pieces of aluminum 6" long, .064, 2 SO.

Have you any questions?

EXHIBIT XIV

AM PERFORMANCE EXAMINATION

WELDING

ALUMINUM BUTT WELD

EXAMINER INSTRUCTIONS

TASK — This is a test of the examinee's ability to make an aluminum butt weld.

GENERAL DIRECTIONS TO EXAMINER

1. Read very carefully and thoroughly the "Examinee Instructions." Be sure you understand the test and the method for administering it before attempting to give it to the examinee.
2. Assign the examinee to a work station and provide him with the tools and material.
3. Have the examinee read the "Examinee Instructions," then read them out loud to him. Ask him if he has any questions. If he does not understand any part of his directions, re-read that part out loud to him until he understands what he is to do. Supply no accessory information. Remember, this is a test situation, not an instructional one. Make sure the examinee understands what he is to do before he starts.
4. Have the examinee secure the work station when he is finished.

SCORING — Carefully observe the examinee as he works. If you think that the examinee should receive credit on an item do nothing to the number which appears after the observation called for on that item. If no credit is to be allowed on an item, encircle or X over the number which appears after that item. Do as much of the scoring as possible at the time that a particular sub-operation is performed. Do not rely on your memory. Give complete attention to the scoring. Of course, the measurements of the final product will be made after the task is completed. To perform these measurements, first measure in 3" from the start of the weld. Make a mark at this point. Then make the measurements and score those observations called for which pertain to the first half of the final weld. On finishing those measurements and observations which pertain to the first half of the final weld, make those which pertain to the second half. The total score is the sum of the non-encircled or non-Xed credits. On completion of the test, add up the non-encircled or non-Xed credits and enter the total in the space provided.

TIME — Time is counted as soon as the examinee understands what he is to do and is told to start. Record the time that the examinee starts and finishes in the appropriate spaces. For credit on the time item, the examinee must finish in 17 minutes or less. Do not tell the examinee what the time allowance is. If the examinee does not finish in 17 minutes, allow him to finish. He receives no credit, however, on the time item unless he finishes in 17 minutes or less.

EXAMINER'S NOTE:

Make sure that the weld is made lengthwise, e.g., the weld should be 6" long.

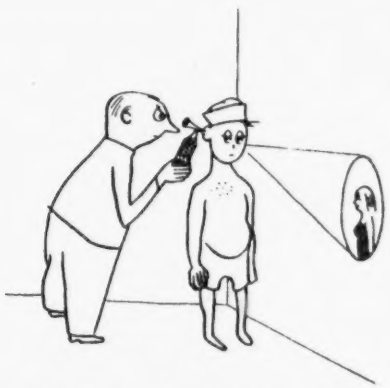
CARE OF EQUIPMENT

CHAPTER V

Examinee Recorded Performance Tests

INTRODUCTION

Some of the tasks Navy men must perform require "information getting" skills. Enginemen, Aviation Motor Machinists, Electrician's Mates, and men in other ratings are frequently called upon to make repairs. Usually in order to make a repair, they must first get some information; they must find out **what** is wrong and they must find out **where** the defect is. If an Engineman finds that the source of difficulty is in the fuel injection system, then he must find out where in that system the difficulty is. Only after he gets this information can he go ahead and make the necessary repairs.



FINDING THE DIFFICULTY

Information-getting abilities of this sort can be tested using examinee recorded performance tests. The examinee is presented with a practical problem and he records his solution on a sheet of paper, or he gets some information as a result of the work he does and records it. The information is then a record of the quantity and quality of the work he did. From it, scores can be obtained which indicate how well he performed.

The applicability of some types of examinee recorded performance tests may be rather limited aboard ship. But, since these tests do not require an examiner for each examinee, they are valuable performance measuring devices where many men must be tested in a short period of time. They may also be widely used with the Craftsman Method of teaching, where it is desirable to test men's ability to **apply** the fundamental principles that are taught.

The information the examinee records can be:

1. **Measurements.** Bearings, distances, resistances, dimensions.
2. **Locations.** The location of a short in a cable, the location of a worn part, the location of a leak in a fuel system.
3. **Interpretations.** The identification of radar target "pips," the meaning of instrument readings, the cause of a rough running engine.

The examinee records this information of these kinds in a variety of ways: by writing down a number or a word, or by placing an X on a diagram or in an answer box.

The general steps in building examinee recorded performance tests are:

1. Devise the test task. This means devising a test task for the examinees to do that will:
   a. require the skill you wish to test, and
   b. result in information that can be recorded on a sheet of paper.

2. Devise a record form on which the examinee can record the information he obtains.

3. Write instructions for the examinee and examiner.

These steps will be covered more fully throughout the remainder of this chapter.

DEVISING THE TEST TASK

Few of the tasks that Navy men perform in their daily routine can be used just as they are as test tasks if the examinee recorded performance test is used. Consequently the test task has to be devised. The test task should:

1. Require the type of performance or skill you wish to test, and

2. Result in information that can be recorded.

Suppose you want to test the ability of electrician's mates to find casualties in a controller. You would insert a realistic fault in the controller and have each of them try to find it. Perhaps it would be a short in the pilot circuit or a blown fuse. The information you could have them record on the prepared test form would be the circuit or part of the controller that was faulty.

In the on-the-job situation, Navy men must apply their skills to a wide variety of tasks and in a wide variety of ways. Therefore the test task should provide the examinees with the opportunity of demonstrating how broadly they can apply their skills.

If the electrician's mates are required to find only **one** casualty in **one** controller, the test would cover only a small portion of their controller trouble-shooting ability. Furthermore, the probability is fairly good that some of the electrician's mates will find the casualty by luck. If, however, each of them has to find one casualty in each of **ten** controllers, the test covers a broader sample of controller trouble-shooting skill, and the probability that they will find all ten casualties by luck alone is very small. Increasing the number of such practical problems that the examinees must solve, increases the value of a test as a measuring device. It will be more reliable and will do a better job of separating the better from the poorer men.

DESIGNING THE RECORDING FORM

Examinee recording forms must be specially designed for each test task, and they differ greatly from one test to another. They should be clearly labeled and lined to provide for simple recording of the information the examinees obtain as a result of their work. In addition, they generally contain the instructions for the examinee.

The way the recording form is designed depends on the test task the examinees must perform, and, because of this, there are no hard and fast rules for its construction. The only consideration which can be stressed is this: The form should provide space for recording all the possible types of information that the examinees may obtain which reflects their performance, and yet it should not give the examinees any clues as to what the correct information is. For example, if an electrician's mate is trying to locate one casualty in each of ten controllers, he could record what he finds on a recording form like the one shown in Exhibit XVI on the following page.

The parts of the controllers in which the casualties might occur are listed vertically on the left and the identifying numbers of the controllers horizontally across the top. The controller is subdivided into a sufficient number of categories (parts or circuits) so that the list offers no clues to the examinee as to where the troubles may be located. Yet the list is sufficiently descriptive so that he can clearly indicate where he thinks the troubles are located.

A recording form for a test of enginemen's ability to interpret instrument panel readings might look like this:

EXHIBIT XV

IN COLUMN I PLACE THE NUMERICAL READING FOUND ON THE INSTRUMENT.

IN COLUMN II INDICATE WITH A "Yes" OR "No" WHETHER OR NOT THE READING IS WITHIN THE NORMAL OPERATING RANGE.

IN COLUMNS III AND IV GIVE THE MAXIMUM AND MINIMUM READINGS AT WHICH IT WOULD STILL BE SAFE TO OPERATE THE ENGINE.

TIME LIMIT: TEN MINUTES.

Instrument Number	I Reading	II Within Norm. Range (Yes or No)	III Maximum Safe	IV Minimum Safe
1.				
2.				
3.				
4.a				
4.b				
5.				

EXHIBIT XVI

CONTROLLER TROUBLE SHOOTING

DIRECTIONS: Each controller is energized. There is a trouble in one of the circuits of each controller. At each controller, push the start button and observe the action of the motor and/or controller. Using the information you get from this operation and the equipment provided locate the troubles in each of the controllers. Indicate your answers in the table below by placing an (X) under the controller number and opposite the condition or part which is causing the trouble.

FOR EXAMPLE: If you found a trouble in the start button of controller "B" you place an (X) as shown below:

PILOT CIRCUIT	Controller Number A	B	C	D	E
a. Control fuses					
b. Stop buttons					
c. Start buttons		x			

The main thing is to find the troubles. Don't waste time, extra credit will be given for rapid work, but **observe all safety precautions.** Are there any questions?

POWER SUPPLY CIRCUIT	CONTROLLER NUMBER 1	2	3	4	5	6	7	8	9	10
a. Blown fuses										
b. Open circuit breaker or switch										
c. Poor connection in circuit										

PILOT CIRCUIT (Control Circuit)	1	2	3	4	5	6	7	8	9	10
a. Control fuses										
b. Stop buttons										
c. Start buttons										
d. Auxiliary contacts										
e. Overload relay contacts										
f. Poor connections										

MAIN LOAD CIRCUIT (Motor Load Circuit)	1	2	3	4	5	6	7	8	9	10
a. Main line contacts										
b. Overload relay elements										
c. Poor connections										

KEEP YOUR ANSWER SHEET FACE DOWN WHEN NOT WRITING

Below is an example of a recording form that could be used to record the bearings a quartermaster takes with a pelorus.

Name:_____ Date:_____ Place:_____

OBJECT	BEARING
1. Radio Beacon	
2. Lighthouse	
3. Smoke Stack	
4. Buoy	
5. Etc.	

Often a mechanical drawing or circuit diagram can be used as a recording sheet. The diagram presented in Exhibit XVII shows one for an electrician's mate circuit trouble-shooting test. The examinee actually does the work on a real circuit that is mounted on a bulkhead. He uses the circuit diagram to record what he finds wrong with the circuit.

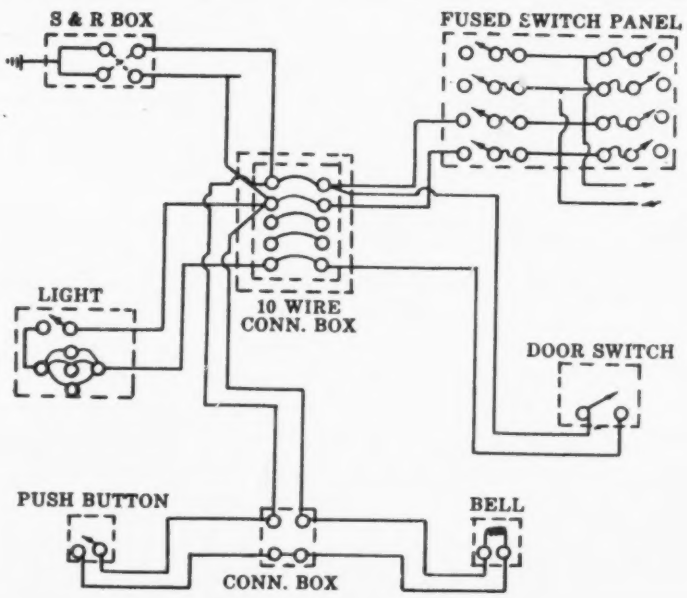
EXHIBIT XVII

CIRCUIT TROUBLE SHOOTING

DIRECTIONS: The circuit is energized. Below is a diagram of the circuit. When given the word begin, you will locate as many shorts, opens, or bad fuses in the circuit as you can. If the circuit is grounded, locate the source or sources of the ground. Indicate on the diagram what faults you find and where they are. You may disconnect any of the wires or remove the fuses. Do not correct any of the faults you find. After time is called you will reassemble the equipment and leave it just as you found it.

Do the work just as if you were aboard ship. Observe all safety precautions and standard procedures. Are there any questions?

TIME LIMIT: Twenty minutes.



The recording form for the circuit trouble-shooting test could also be made like this:

CIRCUIT TROUBLE-SHOOTING

		GOOD	GROUNDED	OPEN	SHORTED
S and R Box					
Light					
Push Button					
Door Switch					
Bell					
Connection Box					
10 Wire Connection Box					
Switch panel					
CABLES	1				
	2				
	3				
	4				
	5				
	6				
	7				

Batteries are a major consideration aboard submarines. A chief aboard a submarmine could test the skill of new strikers by sending them to the battery well with a recording form like the one below and instructions to take gravity, height of electrolyte, and temperature readings.

INDIVIDUAL CELL READING

CELL NO.	GRAVITY	HT. OF ELECTROLYTE	TEMPERATURE
1			
2			
3			
4			

Many tasks that are done day by day in the Navy require men to read and interpret blueprints. Examinee recorded performance tests can be used to test men's ability with this skill. The recording form for a blueprint reading test could contain a list of part names next to which the examinee is required to write the correct dimensions.

INSTRUCTIONS FOR THE EXAMINEE

The examinees' instructions should contain the following information:

1. A statement describing the test task to be performed. For example, "You are to take ground readings on the armatures of these five generators."
2. Instructions on how he is to record his work on the answer sheet with an example. For example:

Using the information you get from this operation and the equipment provided locate the troubles in each of the controllers. Indicate your answers in the table below by placing an (x) under the controll . number and opposite the condition or part which is causing the trouble.

FOR EXAMPLE: If you found a trouble in the start button of controller "B" you place an (x) as shown below:

	Controller Number				
PILOT CIRCUIT	A	B	C	D	E
a. Control fuse.					
b. Stop buttons					
c. Start buttons		x			
d.					

3. A list of the tools, measuring devices, and equipment the examinees will use in performing the test job.
4. The time limit if there is one.
5. If the test job is hazardous, or if there is any danger of the examinees damaging equipment, a precautionary statement should be made in the instructions. For example: "The circuit is energized. To avoid harmful shock observe all safety precautions."

INSTRUCTIONS FOR THE EXAMINER

In many instances the test builder will also be the test administrator, in which case the examiner instructions simply serve to insure that each time the test builder gives the test it will be given in the same way and under the same conditions. If some other person is to administer the test, the instructions should be detailed and complete.

The examiner's instructions should contain:

1. A statement of the examinee's task.
2. A description or list of the tools, machines, and materials he is to provide for the test.
3. Instructions for inserting casualties or other special conditions in the equipment.
4. Instructions on how to score the tests.
5. Instructions on maintaining good testing conditions.
6. Instructions on the time limit if there is one.

SCORING EXAMINEE RECORDED PERFORMANCE TESTS

A point that is obvious yet often overlooked is that in order to score the information which the examinees record, you yourself must have the correct information. In devising the test job, consideration must be given to this fact. The examiner must be able to get the correct information either before or after the examinees take the test.

There are many ways information-getting practical performance tests may be scored depending on the nature of the task performed and the type of information obtained. Generally speaking, such tests are scored on the basis of the correctness of the information that is recorded.

1. **Number of Correct Responses.** One way to score performance tests is to assign one point for each correct measurement, location, piece of information, interpretation, or name. If there are ten troubles in a circuit the examinee is required to analyze, he may be assigned one point for each one he finds. If

some are clearly more difficult to find than others, extra points may be given for finding the more difficult ones, and additional points may be given for correctly indicating what the trouble is. Also, points can be deducted for finding casualties which do not exist. This is discussed in more detail under "errors."

In the case of measurements, one point can be given for each one that is correct or within specified tolerances. Care must be taken in scoring measurements not to require measurements which are beyond the precision of the measuring devices the examinee uses. For example, you couldn't expect the examinee to make measurements accurate to a thousandth of an inch if his measuring device measures only in tenths. Care must also be taken to base scores on the differences in the work the men do and not the differences that result from the equipment they are using. Machinists are familiar with the fact that it is often impossible to turn out parts to the same degree of precision with two supposedly identical lathes. (See Chapter VI, "Administration of Performance Tests," for additional information concerning differences between tools and machines.)

2. **Errors.** In many cases deducting points for errors in the examinee's answers results in an important change in the score given for correct information. In some cases it is important to score both the correct and the incorrect information. For example, if the examinee is required to locate casualties in an engine, he may be given one point for each one he finds. His score would be the total number of casualties he locates. But, suppose two men each find five casualties, and one of the men reports two additional casualties that do not really exist. These are errors and the man who found casualties which did not actually exist should receive fewer points than the man who found the five casualties without making any errors. In this case points should be deducted for the errors.

POINTS SHOULD BE DEDUCTED FOR ERRORS

3. **Time.** The time required to do a job is a measure that should be used when:

   a. Time is an important part of the job, or

   b. The quality of two men's performance is nearly equal. In this case the faster man may be considered more proficient in performing the task.

SCORING AIDS

When large numbers of tests are to be scored, a template can be devised which when placed over the recording form will show if the examinee has placed X's in the proper boxes. The template is prepared by cutting a series of windows in a piece of cardboard. The windows are cut so that if the edges of the template are placed squarely over the edge of the examinee's record form, the boxes in which the correct answers are placed show through the windows. The total score of an examinee is obtained by simply adding the X's that appear in the template's windows.

SUMMARY

There are many tasks which require information-getting skills. Trouble-shooting, measuring, interpretation, and identification skills are of this type. Examinee recorded practical performance tests can be used to test these skills.

The steps in building these tests are as follows:

1. Devise a test task for the examinees to perform. This task must require the skill you wish to test, and its completion must result in information which can be recorded on a sheet of paper and which indicates the quantity and quality of the work the examinees do.
2. Design a recording form on which this information can be recorded by some method such as writing a number or a word, or by placing an X in a space or square.
3. Write instructions for the examinees and examiners.
4. Determine how the performance can be scored. There are three ways answer sheet tests can be scored: by giving one point for each correct response (casualty found, correct measurement, proper interpretation, correct identification, etc.), by counting one point off for each error, or by measuring the amount of time it takes to perform the test job.

CHAPTER VI

Using Performance Tests

TEST ADMINISTRATION

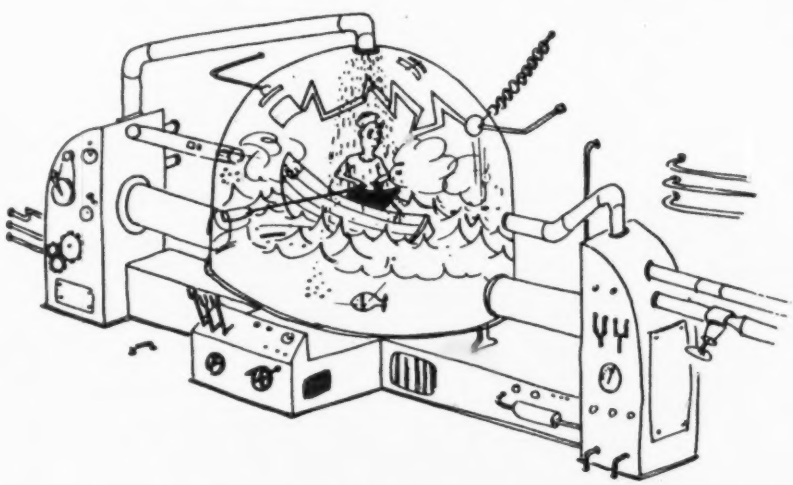
A considerable amount of time and effort is involved (1) in properly selecting the tasks and skills around which performance tests should be constructed, (2) in translating these into test tasks, (3) in devising record forms and (4) determining satisfactory ways of scoring the record of performance that results. All of this effort will be largely wasted, however, if the tests are not administered under proper conditions. There are two general rules which universally apply to test administration:

1. **Best possible conditions.** Tests should be given under the best possible conditions so that all men taking the tests have an opportunity of doing their best work.

2. **Uniformity.** Each time a test is administered it should be done in the same way and under the same conditions. The scores made by different men on a test have no basis for comparison unless all examinees have the same chance to perform well.

TESTING AREA

It would be highly desirable if a regularly assigned space could be set aside for conducting performance tests. This is more likely to be possible in a school situation or on a Navy base than it is aboard ship. However, whatever the situation, a space which is as comfortable and well-lighted and ventilated as possible should be selected. Distracting influences should be held to a minimum during testing periods. If possible this means that testing should be scheduled when other ship's work is not highly noisy or otherwise distracting. Examinees who are waiting to be tested, or interested outside observers, should not be permitted in the testing area unless they are somehow connected with the administration of the tests.



KEEPING CONDITIONS UNIFORM

The thing being sought here is uniform conditions for all examinees. It is realized that in the operating situation it is not always possible to achieve such conditions, but to the extent that they can be achieved the more meaningful the resulting scores will be.

TEST EQUIPMENT

The equipment the examinees use should be in the same condition from one test administration to the next. If it is not, it is not possible to know to what extent scores made by different men result in part from different conditions of the equipment.

For example, if one man is tested using sharp, clean, and well operating tools, and later, another man uses dull, dirty tools, the difference between their scores may in part be due to the fact that one man had better equipment to work with than did the other. The same thing is true, of course, of materials. If a man is supplied with some materials from which he is to construct something, then all men so tested should be provided with materials of the same kind and quality.

If the tests happen to be of the trouble-shooting variety, the casualties which are inserted in the equipment should be the same each time the test is administered. There is one possible exception to this last statement. It may be desirable to deliberately change the nature of the casualty, or its location, from one testing period to another because of the fact that in time the word gets around among the men as to just what kind of casualty to look for. In this case it is desirable to introduce different casualties, but great care must be exercised so that any alternative casualty which is introduced is equally as hard to handle as the original one. In other words, it would not be any easier or any harder for men to find an alternative casualty than the original one.

There is one other comment that should be made concerning testing equipment. That is, the machines used in testing may affect the scores. This is particularly true if a number of men are given the same task to perform using different machines.

For example, in a shop, the task may be to turn out a shaft on a lathe, with five men to be tested at one time, each man using a different lathe. In this situation part of the differences in the final products may result from the fact that the five lathes themselves are different. Obviously it would not be fair to penalize a man who had used a lathe which was not operating as well as some of the others. The one way to be sure of avoiding this problem is to have each man use the **same** piece of equipment in turning out his product. This will be inconvenient if you may want to test a number of men at one time. If several machines are used, then some effort must be made to find out from those people who are very familiar with the gear in question which ones perform in nearly identical fashion. Then the test should be restricted to using these particular equipments.

If there is enough time, you can have all of the examinees turn out one shaft using each lathe. Each man's score would then be the average score of all the shafts he turns out.

TIME

If it is considered desirable to have a time limit on a given test, it should be the same for every test administration. Furthermore, if credit is to be given for completing the task within certain specified time limits, this should be made clear to the examinee and these time limits should be constant from one test administration to the next.

TEST INSTRUCTIONS

Test instructions should be very clearly written and both the examiner and examinee should have written copies. Before the test begins the examinees should have a chance to look at the instruction sheet and to read it over. Afterward, the examiner should also read the instructions aloud while the examinees listen and follow on their own copies. When the examiner is finished, he should ask if there are any questions relative to the task and the test setup and should take care that each examinee clearly understands what he is to do.

The examinee's instruction sheet should indicate the following:

1. Exactly what job is to be accomplished.
2. The equipment and tools which are supplied or which may be used if the examinee wants them.
3. The general features of the performance that are to be scored. This means that the examinee should be told if he's to be scored on accuracy, safety precautions, care of equipment, speed, etc. He should not be told the specific score points, but may be told whether the emphasis in the scoring is on speed, accuracy, safety, etc.
4. The time limit if there is one.

In addition to examinee instructions, a sheet of instructions also should be devised for the examiner

This is desirable even if the person who constructs the test expects also to be the examiner. The instructions help the examiner administer the test the same way each time he gives it. And, in addition, it may be that in the future some other individual will administer the test. He should know exactly how the test situations were set up so that he can duplicate the original test in as many respects as possible.

Generally speaking, one of the most important things that you can do to assure standard test conditions is to devise clearly understood examiner instructions. To the extent that you do this, you will ensure that each examinee will have his best opportunity to perform and that the resulting scores will be comparable for different administrations and different examiners.

PRE-TESTING THE TEST

It is very often difficult to determine just how good a test is until it has been tried out. Often a number of problems arise which simply cannot be anticipated until somebody actually starts to perform. Therefore, before the test becomes generally used, a pre-test should be conducted on a small group of men to determine the answers to some of these questions:

1. Do the instructions cover all the necessary points? Are they understandable?
2. How much time does it take to perform the test? Is the time limit too long or too short?
3. Do the men have the proper equipment and tools or should something be added?
4. How difficult is the test for the persons who will be taking it? Is it too difficult or too easy?
5. Does the test appear to be a good measure of men's ability to do the task around which it has been constructed? Does it really measure the skill you set out to measure?
6. If the test is an examinee recorded type, is the recording form understandable to the examinees and easily used?
7. If the test is of the check list type, are all the essential steps included on the check list? Are any of the steps which have been included not essential for evaluating performance?
8. If it is a final product type test, are there sufficient features in the final product so that the differences in men's skill will be properly reflected?

The examinees on the pre-test should not be the same men who will eventually be evaluated on the test itself. Sometimes helpful suggestions for the improvement of the test can be obtained by pre-testing the test on men whom you know to be experienced in performing such a task. However it must be remembered that if the examinees are strikers, the performance of the second or first class petty officers on the pre-test will not be indicative of what you will get from strikers. A test may be very difficult for strikers and yet be easy for rated petty officers. The other possibility exists, of course, that the test may prove to be of moderate difficulty for rated men and will not be greatly more difficult for strikers. Some recent studies have shown that as far as some practical factors are concerned, strikers and third class petty officers frequently perform as well or nearly as well as some of the higher rated men. Perhaps the best rule is that the pre-test should be carried out using examinees who are as nearly alike as possible to the men who will eventually take the tests.

After the test has been pre-tested and the questions concerning it have been answered it will be ready for use.

MAKING A FREQUENCY DISTRIBUTION OF SCORES*

Once all of the tests for a group of examinees have been scored, the first step in analyzing those scores is to put them into a distribution. Such a distribution is a way of arranging scores so that more meaning can be derived from them. It will help you to answer three questions:

1. **Did the test discriminate?** That is, did all the examinees tend to make about the same score or was there a wide variety of scores obtained on the test? Can it clearly be said who was the best performer, who was the next best, who was least best, and so forth?

2. **How difficult was the test?** The frequency distribution of scores will reveal whether or not the entire group tends to get perfect or near perfect scores, in which case the test may be regarded as too

* See NAVPERS 16808A, CONSTRUCTING AND USING ACHIEVEMENT TESTS, for additional information concerning such topics as converting raw scores to Navy grades, establishing passing grades, and test administration.

easy, or it may reveal that most of the examinees got very low scores, in which case the test might be judged too difficult.

The third possibility, of course, is that the examinee's scores fell pretty evenly over a wide range with some people getting high scores and some getting low and the majority in the middle. This last arrangement of scores is generally the one that is desired and reflects a proper difficulty level.

3. **Does the test do what you want it to do?** Including the two previous points and in addition to them is the consideration of whether or not your test does its job. A score distribution and a close look at your individual test results will tell you this. For example, suppose you want to test a man who has just come aboard to see if he can perform well enough on a control cubicle to stand a watch. You would be best advised to test your present men and look at your score distribution to see if it separates the qualified watch standers from the men who are not qualified. Your test may not tell you who is best among the qualified men, but if it separates the qualified from the unqualified, then you can test your new man with it and see how he compares with the two groups. It does what you want it to do.

To construct a score distribution, all the possible scores from highest down to lowest are first listed in a vertical column. Then for each score obtained by a man on the test a tally mark is made opposite that score in the appropriate place as shown in Exhibit XVIII below. Notice in this example that one man made a score of 8, two men made a score of 7, four men made a score of 5, and so forth. At the bottom of the distribution the total number of men who have taken the test thus far is indicated. In this case, that number is 15.

EXHIBIT XVIII

FREQUENCY DISTRIBUTION OF SCORES

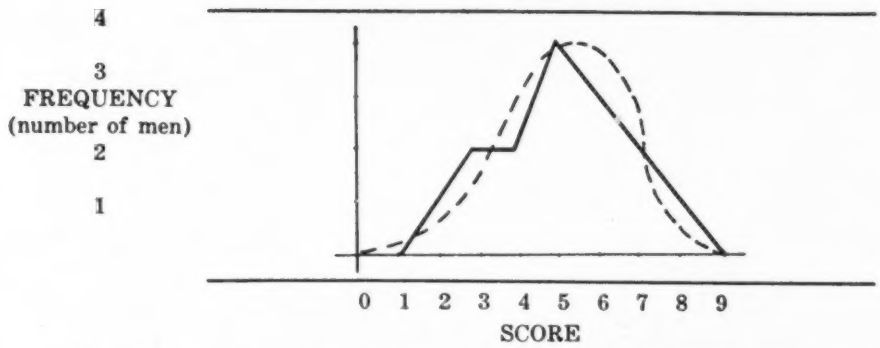
SCORE	FREQUENCY	
	TALLIES	N
8	I	1
7	II	2
6	III	3
5	IIII	4
4	II	2
3	II	2
2	I	1

N = 15

In order to get a picture of what the frequency distribution looks like and to reveal more clearly the way our test is performing, it is common procedure to convert a frequency distribution into a curve like the one shown in Figure 1. This is done by listing the raw scores from the test along the horizontal axis at the bottom of a graph and the number of men making each score along the vertical axis.

FIGURE 1

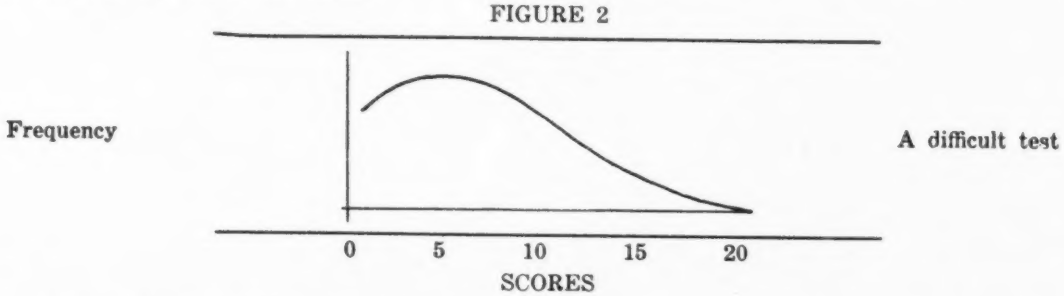
CURVE OF SCORE DISTRIBUTION IN EXHIBIT I



When only a few cases have been assembled on each test the corresponding curve can be expected to be quite rough and jagged in spots. However, the larger the number of cases accumulated the smoother it will become, so that in time it actually takes on the characteristics of a smooth curve such as that shown by the dotted line in Figure 1. The curve makes it easier for you to picture the distribution of scores on a test, and thus helps answer the two important questions about the difficulty level and degree of discrimination associated with the test.

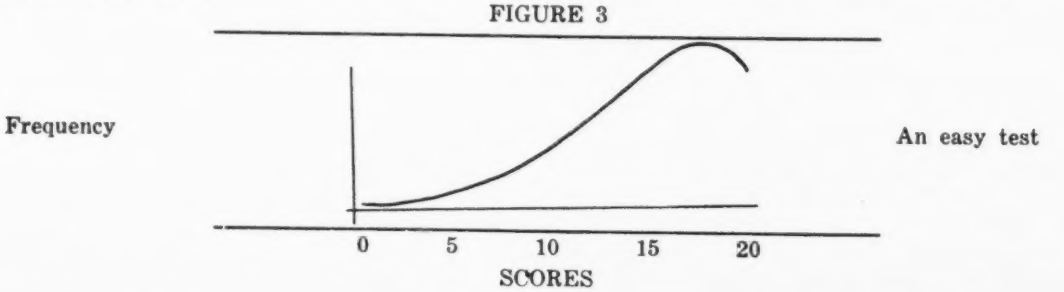
Three frequency curves are shown on the following pages that might have been obtained from three different tests. Notice Figure 2. This curve indicates that the majority of the people did not do very well in the test since most of them got low scores. If the test was properly constructed, this may simply mean that the test was too hard for the examinees who were tested. On the other hand, the possibility always remains that this may be an area of weakness for the examinees and that special training is in order. But whatever the interpretation, the indication we get from the curve is that this was a difficult test for the men who have been tested.

FIGURE 2



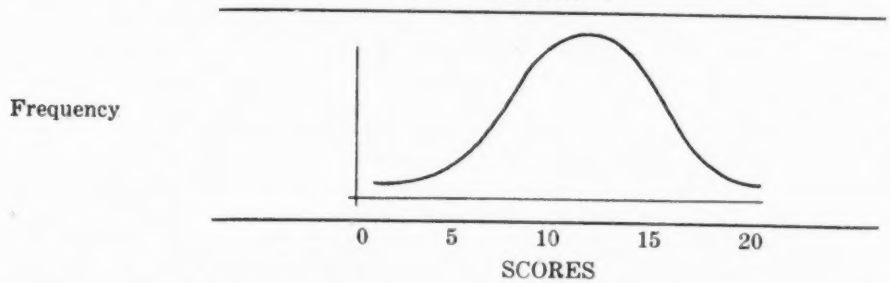
In Figure 3 we see the opposite condition holding. Very few men received low scores and the majority of the men made perfect or near perfect scores. This is a test that was relatively easy for the men who were tested, and it should be made more difficult if we want to discriminate among the several men who got the same scores.

FIGURE 3



Finally in Figure 4 we have a curve which does not reveal a bunching up at either end of the scale but rather, if there is any bunching at all, it is near the middle of the scale. Generally speaking, this kind of curve describes a test task which is at the proper difficulty level for the men being tested. We have neither a lot of men failing the test nor a lot of men making perfect scores. There are a few in those categories, but the majority of men made scores somewhere near the middle.

FIGURE 4



A test of proper difficulty level for maximum discrimination

The matter of discrimination can also be answered with reference to the curves of scores. The curve in Figure 4 describes the test that is achieving the greatest amount of discrimination of the three. This is true because there is the least amount of bunching of scores. In Figure 2 we have a lot of people making scores of zero, 1, and 2. We must necessarily assume that there isn't much difference in the ability of these people to perform or else that our test is not properly discriminating among them. Usually the latter conclusion is correct. A test pitched at a more appropriate difficulty level would reveal considerable differences in ability which previously were obscured because the test was too difficult.

Figure 3 indicates the reverse effect with a large number of men making near perfect scores. On the basis of their scores we again do not have much discrimination among them. If we were to increase the difficulty level of this test we would probably find that the large group of men who made high scores on this test would spread themselves out over a wider range of scores and thus greater discrimination would be achieved.

The fundamental principle of testing that is being considered here is that there are differences in men's abilities. Your job in testing is to obtain scores that indicate these differences. It will help you do a better job if you get scores for each test on a group of men, set up a frequency distribution, and draw a frequency curve. When this is done, much information should be available about whether or not the test is pitched at the proper difficulty level and whether or not you are getting the discrimination that you desire from the test. Generally it will tell you if your test is doing what you want it to do.

ESTABLISHING MINIMUM PASSING SCORES

There are no hard and fast rules for setting the point on a test that is passing.

One suggestion has been made previously in this chapter. This was to give the test to men you know to be qualified and to other men who are not qualified on the particular task and then compare their scores and choose the point which best separates the qualified men from the unqualified.

The basic problem here, however, is a little more complex. It boils down to the fact that if you are giving a test to a group of people you will end up with a list of scores. Each man will have a number which will represent how well he did on the test. So far so good. Now the problem is to determine which number represents **passing performance.** When you think it over you will note that the decision of setting the passing score depends upon the test (how difficult is it?) and upon the men taking it (how experienced are they?).

Your best means of setting a standard is to use your judgment plus that of other people who are proficient in the task. It may be that some of the mistakes a man made will be more revealing than the score itself. Let us say that you have a check list type test and the man made five errors. The score "5" is less important than the particular errors he made. For instance, if you find that all of the errors had to do with care of equipment and did not occur in the part of the test concerning the performance itself, this information may help you arrive at appropriate conclusions.

A note is in order concerning the making of individual evaluations of this sort. It is the aim of performance testing to set up objective standards by which to judge men. If you are asking yourself and others to judge a man's work, then every effort should be made to keep the judgment confined to the work. That is, when you go to others to ask them to evaluate a piece of work, it is desirable to keep the identity of the man out of it. Let the others judge the work and not know who the man is who did it, if this is possible.

Now there are some pointers to be followed in making the judgments as to what is a passing score, besides just inspecting the work for errors and making a judgment on the basis of those errors alone.

The lowest passing score on a final product performance test could be the number of points assigned to a final product that just barely meets the acceptable standards. The acceptable standards for a given product can only be established through the technical judgment of men who are experienced in the test task. The minimum passing score on a check list performance test could be the sum of points given for successful completion of just enough steps to get the job done but not do it well. Here again expert judgment is required.

A remark probably should be made about the need for setting a **realistic** point below which performance is considered as failing. Sometimes there is a tendency to set passing scores so low that everyone passes. This is natural because no one likes to see a man fail. However, it is better for the examinee, in the long run, as well as for the effectiveness of your ship or your training program, to indicate a man's performance as failing if he really does not come up to acceptable standards. Motivation can only be maintained, and learning can continue to progress, only if the trainee has a **reasonably difficult standard** to shoot at, so that once he meets this standard he feels he has really achieved something. Nothing is worse for a testing program than to have standards which are so low or scoring procedures which are so loose that everybody passes even though they may have only half tried. In the long run the men who are the better performers will appreciate the fact that their skill is really being challenged and that they must learn to do something before they can pass the test.

On the other hand, the passing score can be set so high that no one achieves it. This too is unrealistic and tends to discourage men.

If you find after a considerable number of men have been tested, that everyone is passing a given test, you might examine whether the test is too easy, in which case the frequency curve should reveal the answer, or whether you have perhaps set a minimum passing mark which is too lenient. In the former case the difficulty level of the test might appropriately be increased, and in the latter case a more realistic passing score might be adopted.

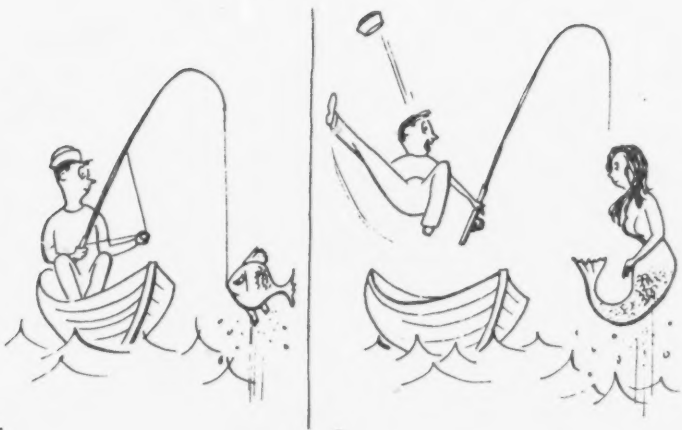
A TESTING TEAM

When practical performance testing becomes a regular procedure in Navy schools and operating units, it is highly desirable to set up a team of experienced personnel who will serve as an evaluation board and be responsible for all phases of the performance test development and administration. You will find, once performance testing gets a good start in your unit, that some of the men will become very much interested in and enthusiastic about it. The testing team should be drawn from these men, and their team responsibilities would be as follows:

1. **To construct performance tests.** The team should work with the other responsible personnel in constructing performance tests.

2. **To organize and schedule testing.** The testing team can serve as a sort of examining board which properly evaluates each man's qualifications for the practical factors.

3. **To analyze test scores.** A testing team should analyze test scores and continually develop the program in the light of the results.



1 2

NOT ONLY INTEREST — BUT ENTHUSIASM

PERFORMANCE TESTING AS A CONTINUAL PROCESS

A really effective testing program, whether it be one of written achievement tests or practical performance tests, necessarily is a continuing one. That is, the test constructor is constantly in the process of revising, changing, and adding new tasks to the test battery. Each time you give a test you will learn something about the test as well as something about your men's performance. You will find some tests too easy and some too hard. Some tests may become obsolete because of equipment changes or changes in procedures. Consequently you will be continually devising new tests and revising old ones. Much depends upon the purposes for which you are testing.

Through all this change you will have an opportunity to build up a file of test tasks designed to sample various practical skills. It is desirable to make this file as complete and as varied as possible. You will find, for instance, that after you have used a test for a while as a final achievement test in practical factors, or as a test of practical factors for advancement in rating, men will begin to do better on it. Their scores become higher.

This may reflect the fact that word has gotten around as to what factors a man will be tested on, and it may be that men are preparing themselves to pass certain specific tasks that they know are coming up. Since you are interested in the general proficiency of men, rather than their ability simply to perform a particular test task, this means that you may have to sample the same skills with different test tasks as time goes on. The ideal way to handle this is to have available a large file of test tasks which will cover all of the practical skills for a given rate. When a man comes up for advancement or when it is time for men to demonstrate final practical performance ability in a Navy school, all the test constructor has to do is to pull a sample of the tasks from the test file, such that the important skills will be represented in their performance. In this way the examinee cannot anticipate the particular task he will be called upon to perform and must prepare himself to be generally proficient in all of the skills required for his rate.

Thus the process of performance testing involves a continuous program of analyzing the skills required by rates, devising test tasks which indicate possession of those skills, building tests, obtaining records, analyzing scores, and revising tests. It is a creative process and, in a way, the job is never over. However, there is no other single procedure which will do so much to insure that Navy personnel maintain the high standards of performance demanded for an effective fleet.

☆U. S. GOVERNMENT PRINTING OFFICE: 1971 - 483-625/2